AN ANALYSIS AND FORECAST OF THE SUPPLY OF FIRST TERM ENLISTEES TO THE UNITED STATES MARINE CORPS

Paul Parsons Darling

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

AN ANALYSIS AND FORECAST OF THE SUPPLY
OF FIRST TERM ENLISTEES TO
THE UNITED STATES MARINE CORPS

by

Paul Parsons Darling

March 1979

Thesis Advisor: P. M. Carrick

Approved for public release; distribution unlimited.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) An Analysis and Forecast of the Supply of First Term Enlistees to the United States Marine Corps		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis; March 1979
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Paul Parsons Darling		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE March 1979
		13. NUMBER OF PAGES 81
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Enlistment Forecast
Marine Corps Enlistment
Box-Jenkins model

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Two distinct analytical techniques were used to develop models in order to forecast monthly first term regular enlistments in the United States Marine Corps. A multiple regression model was derived based on its compatibility with a theory of occupational choice, the intuitive appeal of the explanatory variables, the past literature of manpower supply, and the statistical significance of each variable's impact on monthly enlistments. A second model was

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
(Page 1) S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

#20 - ABSTRACT - CONTINUED

developed by applying the Box-Jenkins methodology to the time series of monthly enlistments spanning the period from July 1973 to June 1978. As a further refinement the residuals from the multiple regression equation were treated as an original time series and the Box-Jenkins technique applied to them. Then the two models were combined and forecasts calculated.

An Analysis and Forecast of the Supply of First Term Enlistees to the United States Marine Corps

by

Paul Parsons Darling
Major, United States Marine Corps
B.A., Dartmouth College, 1974

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 1979

ABSTRACT

Two distinct analytical techniques were used to develop models in order to forecast monthly first term regular enlistments in the United States Marine Corps. A multiple regression model was derived based on its compatibility with a theory of occupational choice, the intuitive appeal of the explanatory variables, the past literature of manpower supply, and the statistical significance of each variable's impact on monthly enlistments. A second model was developed by applying the Box-Jenkins methodology to the time series of monthly enlistments spanning the period from July 1973 to June 1978. As a further refinement the residuals from the multiple regression equation were treated as an original time series and the Box-Jenkins technique applied to them. Then the two models were combined and forecasts calculated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENT

The author wishes to express his gratitude to the Director and staff of the Marine Corps Operations Analysis Group, Center for Naval Analyses, for their cooperation in providing the data base for this study. In particular, the suggestions of Mr. William Cralley were most helpful and appreciated.

I. INTRODUCTION

This thesis is concerned with the occupational choices of young men and specifically the choice of enlisting in the Marine Corps. The problem of whether sufficient numbers of young men would choose the military as a career has been of considerable interest since the advent of the All Volunteer Force in early 1973. For its part, the Marine Corps has had analysts at the Center for Naval Analyses studying the determinants of enlistment supply for some time. The author's interest in the subject was stimulated by an "experience tour" of six weeks duration at CNA during the summer of 1978 while one of the manpower supply studies was in progress. The initial plan was to choose a mathematical model of the supply function solely on the basis of a good statistical fit. Subsequently, it became apparent that while the model should indeed reflect any patterns inherent in the enlistment data, an additional criterion for selecting a functional form was desirable. The form chosen should be deduced from, and consistent with, a model of occupational choice as developed in microeconomic theory [Refs. 3, 5, and 12].

A generally accepted, fundamental model of occupational choice considers each young man to be faced with a choice between enlisting or not enlisting. In this dichotomous world each choice involves certain characteristics. These may be thought of as enlistment attributes and non-enlistment

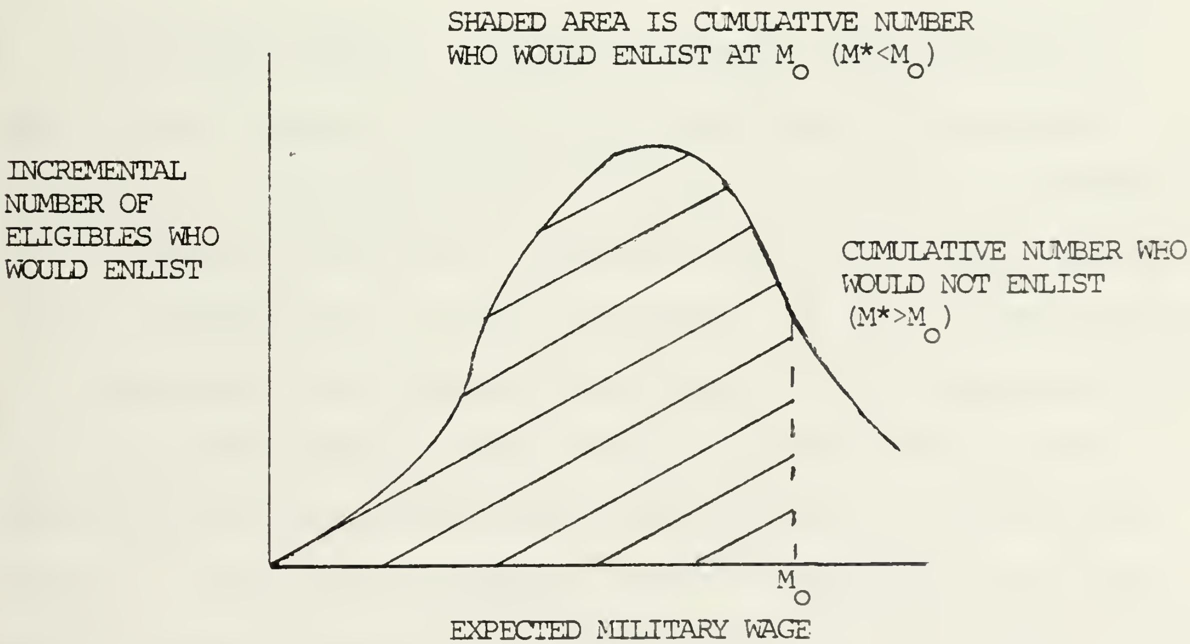
attributes. In principle, each set of attributes may be described in terms of pecuniary and non-pecuniary costs and benefits. The first assumption is that an individual chooses the set of attributes which maximizes his utility function, that is provides him with the most pecuniary and non-pecuniary benefits. It is further assumed that an individual may express these non-pecuniary attributes in pecuniary terms. This means that an individual, faced with enlisting or not is able to put a monetary value on a potpourri of military attributes such as length of initial enlistment; the opportunity for training; location and unit guarantees; benefits available after active service; family separations; deployments overseas; discipline; and arduous working conditions. He then adds this result to the explicit military wage and compares this figure to a similarly derived opportunity cost of his civilian opportunities [Refs. 3 and 8].

Given these assumptions, the potential enlistee may determine a reservation military wage, M*, that would make the sum of the pecuniary and non-pecuniary benefits from enlisting just equal to the sum of pecuniary and non-pecuniary benefits of remaining a civilian. At this wage the potential enlistee would be indifferent between enlistment and remaining a civilian. If the military wage, M, actually offered the applicant exceeded his reservation wage, then he would enlist. If it did not, he would not enlist. Reference 11 provides a comprehensive review of the economics of

job search when the applicant has information concerning the distribution of these wage offers.

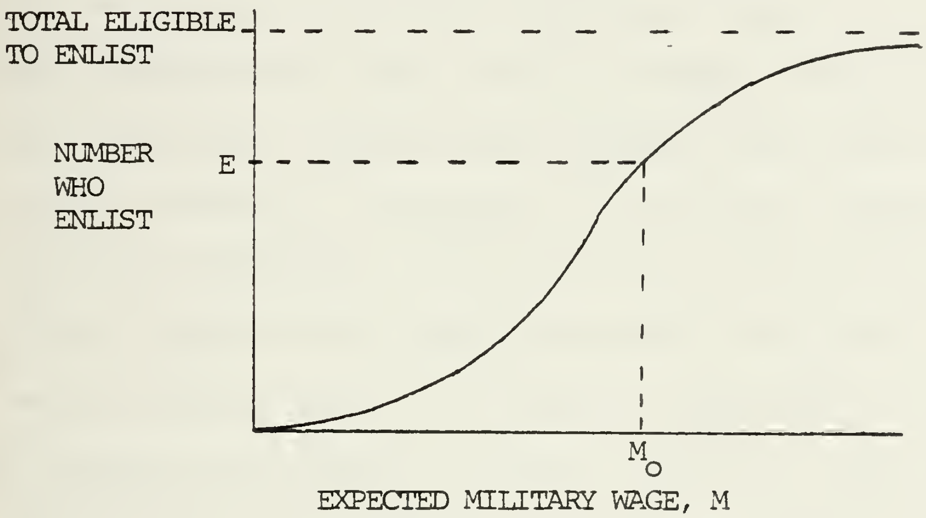
One expects differences in reservation wages among potential enlistees. Such differences are derived from the varying opportunity costs of enlisting and the perceived differences in the non-pecuniary returns to military life, of "tastes" for military service. For example, individuals with excellent civilian opportunities and a strong aversion to military service will, ceteris paribus, have a high reservation wage. Thus, individuals may be arrayed according to their reservation military wages, creating a frequency distribution like Figure 1.

Figure 1



The aggregate supply curve as a function of expected military return (other determinants held constant) may be derived by plotting the number who would enlist at each value of expected military return:

Figure 2



The supply curve in Figure 2 has been heuristically derived for the most obvious tangible aspect of military employment, the expected military wage. The other determinants of the supply of military manpower which were held constant may be categorized in several groups: (1) the dissemination of information to potential recruits, (2) the employment and earnings conditions in the civilian economy, and (3) the population base from which the military draws its recruits [Ref. 3]. The first category concerns recruiting activity, which includes recruiters in the field, mailings, advertising, and logistical support for the recruiting establishment. It

is expected that a larger recruiting effort, other things being equal, would yield more enlistments.

The second category involves those factors that are exogenous to military policy control. As an individual formulates his opportunity cost of enlisting, he considers not only the explicit civilian wage opportunities but also the security associated with those offers as represented by the unemployment rate. As the chances for civilian employment decrease (unemployment rises), more individuals will seek to enlist. Conversely, if civilian earnings opportunities are excellent, fewer enlistees would be expected. The final factor is the population base from which potential enlistees are drawn. If this is expanding one would expect increased enlistments, ceteris paribus. Hence, the above discussion implies that an enlistment supply function may be structured in terms of the factors previously specified:

$$E = f(M,R,C,U,P)$$

where

E = enlistments

M = military wages

R = recruiting effort

C = wages from civilian employment

U = unemployment in civilian labor force

P = relevant population base (e.g., 16-24 year old high school graduate males)

There are two primary techniques for examining the roles that the determinants of the enlistment decision process have historically played: survey instruments and econometric models of enlistment supply. Survey techniques allow the analyst to explore aspects of the enlistment decision process for which other, more "objective" techniques are inadequate due to a lack of data. For example, the reasons affecting an individual's "taste" for military service may be analyzed. However, an econometric model of enlistment supply enables the policy maker to sort out the individual effects of the various factors that contribute to the supply of enlistees. Furthermore, forecasts of future manpower supply levels are possible under a variety of given circumstances. It is the purpose of this paper to model mathematcially the supply function for male non-prior service regular enlistees to the Marine Corps. Two analysis techniques are utilized: multiple regression analysis and the Box-Jenkins methodology. The problem, then, is to model mathematically a suitable cohort of enlistees and to calculate elasticities with respect to the appropriate explanatory variables. The implications of policy decisions and economic factors could then be assessed. Furthermore, the forecasting accuracy of this multiple regression model could be contrasted with that of the model derived by the use of the Box-Jenkins methodology. A review of the more pertinent studies using multiple regression analysis will clarify both the selection of the independent variables and the type of functional form chosen.

II. A REVIEW OF THE LITERATURE ON SUPPLY OF MILITARY VOLUNTEERS

A. THE GATES COMMISSION

The President's Commission on an All Volunteer Force, popularly known as the Gates Commission, involved several studies on the supply of military manpower. Harry J. Gilman stated that one objective of those studies was to estimate the elasticity of manpower supply with respect to military compensation [Ref. 6]. In essence, the studies for the Gates Commission used as a functional form for the supply equation the log-linear model based on time series data for the period 1958 to 1968 (the one cross section analysis utilized data from 1964):

$$\ln E_r = a_o + a_1 \ln (M_p/C_p) + a_2 \ln DP + e + \text{seasonal dummy variables} \quad (2.A.1)$$

where

E_r = E/P, the ratio of enlistments, E, to the eligible population base P

M_p = expected military earnings during length of contract

C_p = expected civilian earnings during same period

DP = some measure of draft pressure

a_i = regression coefficients

e = stochastic error term

A number of functions, including linear, log-linear, log-complement, and logit functions were fitted to the data [Ref. 7 and Table I] and statistical tests for goodness of fit were unable to distinguish amongst them. The coefficient of determination for the time series regressions were in the range from .61 to .82. Therefore, the choice of the model was often due to the author's preference, ease of manipulation, or economic interpretation. In the log-linear model which was ultimately adopted by the Gates Commission, the elasticities with respect to the explanatory variables are constant. This may be shown by proceeding from the basic definition of elasticity as the percentage change in say y for a given percentage change in x:

$$E = \frac{\Delta y/y}{\Delta x/x}$$

and noting that in the limit $\Delta y/y$ is the differential for the logarithm:

$$\frac{dy}{y} = d(\ln y)$$

Rewriting this expression for the elasticity of the enlistment rate with respect to the military to civilian pay ratio one has:

$$E = \frac{d(\ln y)}{d(\ln x)} = \frac{d(\ln Er)}{d(\ln M_p/C_p)} = a_1$$

FUNCTIONAL FORMS FOR SUPPLY EQUATIONS

	LOG-LINEAR	GOMPIT	LOGIT	LOGIT-LOG R	COMPLEMENT	SEMI-LOG
f(p)	$P = aR^{b}$	$P = e^{-AR^{-b}}$	$\frac{P}{1-kP} = e^{a+bR}$	$\frac{P}{1-kP} = aR^{n}$	$1 - P = aR^{-b}$	$e^{P} = aR^{b}$
linear in functional form	$\ln P = a' + b \ln R$	$\ln(-\ln P) = a + b \ln R$	$\ln(\frac{P}{1-kP}) = a + bR$	$\ln(\frac{P}{1-kP}) = a' + b \ln R$	$\ln(1-P) = a' - b \ln R$	$P = a' + b \ln R$
f'	baR^{1-b}	$abe^{-aR^{-b}} R^{-b-1}$	$bP[1-kP]$	$\frac{P}{R} b[1-kP]$	abR^{-b-1}	b/R
ε	b	$-b \ln P$	$bR[1-kP]$	$b(1-kP)$	$b(\frac{1-P}{P})$	b/P
$\frac{d\varepsilon}{dP}$	0	$-b/P$	$\frac{1}{P} - kbR$	$-kb$	$-b/P^2$	$-b/P^2$
weight	ENL/(1-P)	$\frac{\text{ENL}(\ln P)^2}{(1-P)}$	QUAL P(1-kP)	QUAL P(1-kP)	QUAL $\frac{(1-P)}{P}$	$\frac{\text{QUAL}}{P(1-p)}$

NOTE: $a' = \ln a$

TABLE I

from equation (2.A.1). Gilmer justified the constant elasticity by assuming that manpower shortfalls under the All Volunteer Force would be relatively small. Hence, over such a short range the elasticity could be assumed to remain constant. One would only expect a declining elasticity if large manpower deficits were encountered, forcing the services to enlist at the margin people whose preferences obviously lay elsewhere.

The Gates Commission studies made several other general assumptions. The models describing the enlistment decision imply that young men have only two occupational choices: civilian employment or military service [Refs. 2 and 7]. The options of remaining in school or enlisting in another service (for the single service studies) are not included. Additionally, demand restrictions could cause queues for enlistment and result in individuals being turned away. Hence, enlistments for these periods would not represent points on the theoretical supply curve. It was assumed that high school graduates in the higher mental groups were not constrained in this manner [Refs. 2, 6, and 7]. These mental groups were established during World War II to insure a more equitable allocation of new manpower within the Army. However, to extend this concept to all the services, in May of 1951 the Secretary of Defense prescribed identical mental and physical standards for initial induction into any service and assigned service quotas based on an applicants mental category. The categories, listed below,

are still derived from an enlistee's scores on a standardized battery of tests given to all applicants for military service.

Mental Group	Percentile Score
I	93-100
II	65-92
III	31-64
IV	10-30
V	9 and below

B. THE LOG-LINEAR MODEL REVISITED

Three years after the publication of the Gates Commission findings in November of 1970, Jehn and Carroll selected the log-linear model in their study of first-term Navy accessions [Ref. 10]. Their objective was to identify the effects of pay, advertising, and recruiting on new enlistees in the Navy. The model included an attempt to account for more than one career option by adding as explanatory variables other services recruiters and advertising expenditures. The basic model was:

$$\ln\left(\frac{E}{P}\right) = B_0 + B_1 \ln W + B_2 \ln (1-u) + B_3 \ln NA + B_4 \ln NR$$

$$+ B_5 \ln OA + B_6 \ln OR + \text{dummy and seasonal variables}$$

E = enlistments

P = 17-20 year old population of civilian males

W = military to civilian wage ratio

U = unemployment rate

NA = Navy advertising

NR = Navy recruiters

OA = other services advertising

OR = other services recruiters

Jehn and Carroll also limited their enlistment data to those in the upper mental groups in order to insure that the time series of enlistments would represent the "true" supply of volunteers. Furthermore, the authors discounted both military and civilian expected income over a four year period. The employment rate was matched to the cohort of enlistees. Although the study's results were inconclusive (only the coefficient of the unemployment variable was significantly different from zero), Jehn and Carroll did attempt to account for a greater number of influences on the supply of volunteer enlistees to the Navy than had the Gates Commission studies.

C. THE LINEAR MODEL ADVOCATES

When the Gates Commission published it's recommendation to replace the military draft with a purely voluntary system, the role of economic factors in governing the supply of military manpower came under increased analysis. John C. Hause developed a linear model of the supply equation to assess the important impact of changes in employment opportunities on first-term enlistments [Ref. 9]. As in the

previous studies, the author limits his data on enlistments to the higher mental groups. Similarly, he deflates these enlistments by the corresponding population cohort of qualified males. In this manner he eliminates any trend due to cohort size. That is, if tastes for military service are unchanged over the period of analysis and the enlistments from a particular cohort are not demand limited, then the ratio of enlistments to the corresponding cohort size should be constant for given values of the independent variables.

Hause's study differed from the others reviewed here in that his only explanatory variable was the unemployment rate. This is undoubtedly a specification error. His linear model was based on quarterly data from 1957 through 1973. While such restricted casualty precludes meaningful comparisons of elasticities with the other studies, Hause did innovative work by incorporating in his linear model the concept that the probability of an individual enlisting in the service varies with the length of time he has been unemployed. This was accomplished in the following way.

Let $p(\tau)$ be the conditional probability of enlisting given the applicant has been unemployed for a period of length τ. If the fraction (density function) of the population cohort that has been consecutively unemployed for a period τ is written $f(\tau)$, then the expected enlistment rate from this source at the current point in time t is:

$$e(t) = \int_{t-t_o}^{t} p(t-\tau)\, f(\tau)\, d\tau$$

where t_o is the longest consecutive length of time any currently unemployed person has been unemployed. Now if $p(\tau)$ is constant over time then the unemployment rate is:

$$u(t) = \int_{t-t_o}^{t} f(\tau)\, d\tau$$

The enlistment rate would be proportional to the unemployment rate:

$$e(t) = c_1\, u(t)$$

where

$$p(\tau) = c_1 \quad \text{a constant.}$$

The assumption that the conditional probability of enlisting is independent of the length of unemployment is not particularly plausible. Hence, Hause assumes $p(\tau) = c_2\tau$ from which it follows that the expected enlistment rate is:

$$e(t) = c_2\, u(t)\, \overline{T}$$

where $\overline{T}$ is the average duration of unemployment for the population cohort generating the enlistments at time t. Using a linear model including seasonal dummy variables, Hause obtained R^2 values around .86. However, when one excludes the seasonal dummy variables, the consequent

conditional coefficient of determination was about .41 when either "unemployment" variable was included in the regression.

Alan E. Fechter developed a linear model of Army enlistment behavior during the Vietnamese Conflict [Ref. 4]. He utilized the following regression equation:

$$E = B_0 + B_1 DP + B_2 VN + B_i X_{it} + e$$

where

E = enlistment rate

DP = measure of draft pressure

VN = index of Vietnamese Conflict based on number of casualties

X_i = explanatory variables such as the military to civilian pay ratio and unemployment

Quarterly enlistment results during the period 1958 to 1968 were analyzed. In concert with previous studies, Fechter limited the enlistment data to white males in the top three mental categories. These figures were then deflated by the eligible white male civilian population, aged 17 to 20. Military pay was defined as the sum over the period of enlistment of cash pay plus in-kind benefits all discounted at 30%. Civilian pay included the earnings of persons who were comparable to the enlistees in selected characteristics, such as age, race, and education. It was also appropriately discounted. Fechter's study included linear models with a pay

ratio as an explanatory variable and those with both military and civilian pay as separate explanatory variables. The pay elasticites for the latter linear models were consistently greater than those for the relative pay models. Both absolute and relative pay models had R^2 values of about .71.

D. THE LOGIT MODEL

Richard V.L. Cooper [Ref. 3] chose the logistic econometric model for three reasons: (1) such a model binds the dependent variable between 0 and 1; (2) the resultant supply curve has an S shape; and (3) the relevant elasticities initially increase and then decrease over the range of the independent variables. The supply function has the form:

$$S = \frac{1}{1 + e^{-(B_o + B_1W + B_2R + B_3U + B_4D + a)}}$$

which, with some algebraic manipulation, may be put into linear form:

$$\ln\left[\frac{S}{1 - S}\right] = B_o + B_1W + B_2R + B_3U + B_4D + a \qquad (2.D.1)$$

where

S = enlistment rate

W = wage ratio

R = measure of recruiting effort (number of effective recruiters)

U = unemployment

D = seasonal dummy variable

a = stochastic error term

Following tradition, Cooper restricts his semiannual enlistment data over the period from July 1970 to June 1976 to high school graduates in the top three mental categories. He deflated these figures with a weighted average of the 17 to 21 year old male high school graduate population since age cohorts have different probabilities of enlisting. The unemployment rate was matched with the age group of the enlistees. Military pay for the multiple regression using Marine Corps enlistments was discounted over a period of three years to reflect the most common enlistment. The discount rate used on the Regular Military Compensation figures was 20%. Regular Military Compensation for military personnel includes base pay, quarters and subsistence allowances, plus the tax advantage resulting from the latter two categories of remuneration being tax exempt. In calculating the discounted stream of income over the enlistment period, Cooper assumed that enlistees spend the first four months as E-1's, the next eight months as E-2's, the following year as E-3's, and thereafter as E-4's. Additionally, he assumed the applicant was unmarried for the first two years of his enlistment and married thereafter. Cooper noted that including an independent variable measuring the ratio of the number of Marine recruiters in the field to the number of effective Army

recruiters increased the explanatory powers of equation (2.D.1) by about one third.

The multiple regression equation derived was

$$\ln\left[\frac{S}{1 - S}\right] = -3.727 + 1.649W + 1.715R + 1.648U - 0.162D + 1.648r^*$$

where the variables are as previously defined and r* equals the ratio of Marine to Army recruiters. The coefficient of determination for the Marine enlistment data was in the range from .5 to .7 compared to .7 to .9 for the regressions using Navy and Army enlistment data.

The most comprehensive study of the supply of enlistees to the services was done by D.M. Amey, A.E. Fechter, D.W. Grissmer, and G.P. Sica in June of 1976 [Ref. 1]. After providing an excellent review of previous research in the field, the authors proceeded to analyze quarterly Army enlistment data disaggregated by level of schooling and mental category for the years 1958 through 1972. The authors investigated a variety of models based on the logit supply function:

$$\ln\left(\frac{e_t}{1 - e_t}\right) = B_o + \sum_{i=1}^{n} B_i X_{it} + u_t$$

where

e_t = enlistment rate in quarter t

X_{it} = military and civilian pay, the unemployment rate, probability of being drafted and other explanatory variables to account for seasonal

variations and political crises during the period under study.

u_t = stochastic error term

The models were differentiated by changing the set of independent variables. Relative and absolute pay variables were combined with various seasonality assumptions in the so-called static models. The model was termed dynamic when the enlistment rate lagged one quarter was included as an independent enlstment determinant. The authors ranked the models by the accuracy of their forecasting ability as measured by the lowest root mean squared error of the residuals;

$$\sqrt{\sum_{i=1}^{n} (e_t - \hat{e}_t)^2/(n-1)}$$

where

n = number of observations in time series

e_t = observed enlistment rate

$\hat{e}_t$ = expected enlistment rate based on multiple regression results

The two top-rated models excluded seasonal dummy variables and included military and civilian pay as separate explantory variables vice as a pay ratio. Additionally, the top-rated model included the enlistment rate lagged one quarter as an independent variable whereas the runnerup did not.

The study by Amey et al., also utilized three other basic models to test volunteer response to pay and unemployment changes. The results were based on more recent enlistment time series data (1971-1976) from each of the services. The basic models were:

LINEAR MODEL (LS1):

$$E_t = B_o + \sum_{i=1}^{12} B_i \delta_{it} + B_1 W_k + B_2 U_k$$

MULTIPLICATIVE SEASONAL MODEL (MS1):

$$E_t = \prod_{i=1}^{12} e^{B_i \delta_{ik}} (B_1 W_k + B_2 U_k + B_o)$$

COBB-DOUBLAS MODEL WITH SEASONALS (MS2):

$$E_t = B_o \prod_{i=1}^{12} e^{B_i \delta_{ik}} (W_k^{B_1} U_k^{B_2})$$

E_t = volunteer rates for month t

W_k = pay ratio for month k where k may represent a lag of +6, +4, +2, 0, -2, -4, -6

U_k = unemployment rate for month k where k may represent a lag of +6, +4, +2, 0, -2, -4, -6

B_i = regression coefficients

δ_{ik} = 1 if i = k MOD (12), 0 otherwise

The models were estimated using ordinary least squares techniques. A stepwise procedure was used to determine which lag of the independent variables was the most appropriate. It is not clear from Ref. 1 why the logit model was abandoned in favor of the above three types. However, Cooper [Ref. 3] concluded that all of the models previously discussed gave consistent results in the sense that the regression coefficients had the intuitively appropriate signs. The results for Marine Corps volunteer rates are set forth in Table II.

Table II

RESULTS OF AMEY ET AL. REGRESSIONS*

Dependent Variable	Model	Coefficient of pay ratio (lagged 6 months)	Coefficient of unemployment variable	R^2
HSG MC 1 and MC 2	MS1	.0062	.0008	.678
	LS1	.0058	.0007	.886
	MS2	.5923	1.0981	.898

* all coefficients significant at .01 level

III. THE SUPPLY EQUATION FOR MARINE ENLISTMENTS

A. THE DEVELOPMENT OF THE MULTIPLE REGRESSION MODEL

The brief analysis of occupational choice developed in the introduction resulted in the "S" shaped supply curve of Figure 2. Such a curve is called logistic and may also be derived from the solution to the ordinary differential equation:

$$\frac{dS}{dM} = aS - bS^2$$

where, for example, at military wage M equals M_o, the enlistment rate S equals S_o. The solution to this differential equation has the form:

$$S(M) = aM_o/[bM_o + (a - bM_o)e^{-a(M-M_o)}]$$

which results in the logistic or S shaped curve. This implies that the elasticity of the enlistment rate with respect to military pay will first increase and then decrease as pay is increased, and also indicates the asymptotic hypothesis that the enlistment rate can be made as close to one as desired by simply raising military pay sufficiently.

Thus, the logit functional form fulfills the additional criterion of being consistent with the assumed model of occupational choice. The explanatory variables to be included in the supply equation were chosen by considering their

intuitive appeal, the availability of data, and the previous literature concerning the supply of military manpower.

B. THE VARIABLES

1. The Civilian-Military Pay Ratio: CMPR

As a measure of the relative earnings opportunities in the civilian and military occupations a pay ratio was calculated. The national estimate of weekly earnings of production and non-supervisory workers on industry payrolls was used as a measure of the prevailing civilian wage opportunity for the population cohort of eligible enlistees. This figure was divided by the weekly earnings of an enlisted man in pay grade E-3 with three years of service and a dependent wife. The military wage included the tax advantage resulting from the tax exempt status of the subsistence and quarters allowances. Although a Lance Corporal with over three years service and a wife is not the profile of the average first term enlistee, an applicant considering enlistment in the Marine Corps would probably consider promotional opportunities and the various allowances when making a wage comparison with any civilian offers. Hence, it was considered appropriate to use a higher wage figure than that given initially to a basic recruit. Furthermore, the higher wage is more comparable to what one would calculate by discounting an income stream over the most common length for a first term enlistment.

2. Unemployment: UNEM

The figure used for unemployment was the deseasonalized estimate of the national unemployment rate for persons aged 16 to 19. This statistic was chosen over the national unemployment rate for all persons over 16 years of age because it more accurately reflects the milieu of the teenage job seeker. The relatively small number of monthly enlistments would have no appreciable impact on the unemployment rate used in the multiple regression. This avoids any "feedback" relationship from the dependent variable to the independent variable which could bias the regression results.

3. Monthly Leads From National Print Media: LEADS

The values of this explanatory variable are the monthly number of qualified leads obtained from postcards which were included in national printed media advertisements. A lead is considered qualified if the respondent meets minimum age and educational levels. This variable was chosen to act as a proxy for the impact of advertising on enlistments. A more direct measure of advertising input such as appropriated funds is difficult to use since the actual placement of an ad may bear little relation to when the money was obligated.

4. A Dummy Variable: DJ

A dummy variable only takes on the values 0 or 1 and is designed to account for anomalies in the factors influencing the values of the dependent variable. In this case a dummy variable was assigned the value 1 for December of

1976 and 0 otherwise. This was to correct for the abnormally large number of enlistments in that month due to the expiration of educational benefits under the G.I. Bill on 31 December 1976.

5. <u>The Number Of Marine Recruiters: EFFREC</u>

The values of this variable are the number of Marine recruiters who were actually on duty each month during the period of the time series. They represent actual canvassers and exclude headquarters' staff and those recruiters who were sick or on leave during a given month.

6. <u>The Dependent Variable: LSV</u>

The initial hypothesis was that the number of monthly enlistments would be a function of the above independent variables (displayed in Table III). However, the functional form of this dependence, the logit model, requires a transformation of the dependent variable to achieve a linear form. The nonlinear equation is:

$$S = \frac{1}{1 + e^{-(B_o+B_1 CMPR+B_2 EFFREC+B_3 UNEM+B_4 LEADS+B_5 DJ+a)}} \qquad (3.B.1)$$

where the independent variables are as previously defined and the dependent variable S is an enlistment rate. To insure that the observed number of monthly enlistments was not restricted by Marine Corps policy and was therefore truly representative of points on the supply curve, the analysis was only concerned with regular enlistments of male high

TABLE III

CASE-N	DATE	EFFREC	VSMART	CMPR	LEADS	DJ	UNEM
1	[illegible]	[illegible]	781.	1.11	0.	0.	[illegible]
2	[illegible]	[illegible]	[illegible]	1.11	0.	0.	[illegible]
3	[illegible]	[illegible]	[illegible]	[illegible]	0.	0.	[illegible]
4	[illegible]	[illegible]	[illegible]	[illegible]	0.	0.	[illegible]
5	7311.	[illegible]	[illegible]	1.07	0.	0.	[illegible]
6	7312.	[illegible]	[illegible]	[illegible]	0.	0.	[illegible]
7	7401.	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
11	[illegible]	[illegible]	[illegible]	[illegible]	0.	0.	[illegible]
12	7406.	[illegible]	[illegible]	1.13	0.	0.	13.0
13	7407.	[illegible]	857.	1.14	0.	0.	[illegible]
14	7408.	[illegible]	1035.	1.14	391.	0.	13.7
15	7409.	[illegible]	1073.	1.16	6797.	0.	[illegible]
16	7410.	1032.	749.	1.05	[illegible]	0.	[illegible]
17	7411.	[illegible]	475.	1.04	11275.	0.	17.0
18	7412.	[illegible]	469.	1.05	[illegible]	0.	[illegible]
19	7501.	[illegible]	[illegible]	[illegible]	[illegible]	0.	[illegible]
20	7502.	[illegible]	764.	1.04	[illegible]	0.	[illegible]
21	7503.	[illegible]	746.	1.04	[illegible]	0.	20.6
22	7504.	[illegible]	[illegible]	1.05	[illegible]	0.	[illegible]
23	7505.	[illegible]	947.	[illegible]	[illegible]	0.	[illegible]
24	7506.	[illegible]	2047.	1.08	[illegible]	0.	[illegible]
25	7507.	[illegible]	[illegible]	1.09	0.	0.	[illegible]
26	7508.	[illegible]	[illegible]	1.10	[illegible]	0.	[illegible]
27	7509.	[illegible]	[illegible]	[illegible]	7325.	0.	[illegible]
28	7510.	[illegible]	1391.	1.06	4317.	0.	[illegible]
29	7511.	[illegible]	[illegible]	1.08	[illegible]	0.	[illegible]
30	7512.	[illegible]	[illegible]	1.07	[illegible]	0.	15.6
31	7601.	[illegible]	1204.	1.06	2777.	0.	[illegible]
32	7602.	[illegible]	1019.	1.07	[illegible]	0.	[illegible]
33	7603.	[illegible]	951.	1.07	10047.	0.	[illegible]
34	7604.	[illegible]	742.	1.07	[illegible]	0.	[illegible]
35	7605.	1772.	[illegible]	1.09	[illegible]	0.	[illegible]
36	7606.	[illegible]	1202.	1.10	[illegible]	0.	[illegible]
37	7607.	[illegible]	[illegible]	[illegible]	[illegible]	0.	[illegible]
38	7608.	[illegible]	[illegible]	1.13	[illegible]	0.	[illegible]
39	7609.	[illegible]	[illegible]	1.13	[illegible]	0.	[illegible]
40	7610.	[illegible]	723.	1.08	[illegible]	0.	[illegible]
41	7611.	1776.	777.	1.08	[illegible]	0.	[illegible]
42	7612.	[illegible]	1302.	1.10	[illegible]	0.	[illegible]
43	7701.	[illegible]	663.	1.07	20415.	0.	[illegible]
44	7702.	[illegible]	[illegible]	1.10	9081.	0.	[illegible]
45	7703.	[illegible]	[illegible]	1.10	[illegible]	0.	[illegible]
46	7704.	[illegible]	[illegible]	1.11	[illegible]	0.	[illegible]
47	7705.	[illegible]	411.	1.12	[illegible]	0.	[illegible]
48	[illegible]	[illegible]	[illegible]	[illegible]	2720.	0.	[illegible]
49	7707.	[illegible]	[illegible]	1.15	0.	0.	[illegible]
50	7708.	[illegible]	613.	[illegible]	[illegible]	0.	[illegible]
51	7709.	[illegible]	590.	1.16	[illegible]	0.	[illegible]
52	7710.	[illegible]	464.	1.07	[illegible]	0.	[illegible]
53	[illegible]	[illegible]	470.	1.07	[illegible]	0.	[illegible]
54	7712.	[illegible]	[illegible]	1.08	[illegible]	0.	[illegible]
55	7801.	1912.	487.	1.09	13671.	0.	15.0
56	7802.	[illegible]	[illegible]	[illegible]	[illegible]	0.	[illegible]
57	7803.	[illegible]	455.	1.08	9785.	0.	[illegible]
58	7804.	[illegible]	[illegible]	1.09	[illegible]	0.	[illegible]
59	7805.	[illegible]	[illegible]	1.10	5511.	0.	[illegible]
60	7806.	[illegible]	590.	1.12	1432.	0.	[illegible]

school graduates in the top two mental categories (variable VSMART, Table III). The assumption was that possible enlistments from this cohort were never turned away due to a quota restriction; a fairly safe assumption for the Marine Corps. Thus the enlistment rate was derived by dividing the monthly enlistments by a corresponding population cohort consisting of male high school graduates, aged 16 to 24, not enrolled in college. This eliminated any trend due to the size of the eligible population for enlistments. If tastes for military service are unchanged over the period of analysis and the enlistments from a particular cohort are not demand limited, then the ratio of enlistments to the corresponding cohort size should be constant for given values of the independent variables.

When the logit function is transformed to linear form so that the technique of multiple regression may be applied, it becomes:

$$LSV = LN\left(\frac{S}{1-S}\right) = B_o + B_1CMPR + B_2EFFREC + B_3UNEM + B_4LEADS + B_5DJ + a \qquad (3.B.2)$$

So the dependent variable used in the multiple regressions was the natural logarithm of the quantity $(\frac{S}{1-S})$ where S equals the enlistment rate calculated using only enlistments by high school graduates in the top two mental categories. These were assumed to be a true indication of the supply

of manpower to the Marine Corps and not demand constrained in any manner.

C. THE REGRESSION RESULTS

In order to properly assess the outcome of the multiple regression analysis, it is necessary to understand the basic assumptions that are made when applying the technique [Ref. 13]. There are four basic assumptions, the first of which is that the dependent variable is linearly related to each of the independent variables. If the relationships are not linear then the methodology of regression analysis cannot be accurately applied to the problem. Although the logit model has a non-linear form initially (equation 3.B.1), with a transformation and a little algebra it may be put in linear form (equation 3.B.2).

The second basic assumption in regression analysis is that of constant variance of the random errors. This assumption is simply that the variance of the residuals remains constant over the range of the time series. A lack of constant variance may be recognized by a pattern in the residual values. If this problem is not corrected, tests of statistical significance become meaningless. For example, one could not legitimately test to see whether a regression coefficient was significantly different from zero. This could result in the assumption of causality between the dependent variable and an independent variable where none existed.

The third basic assumption of multiple regression is that the residuals are independent of one another. Any given residual should not be a function of the error terms preceeding or following it in the time series. The existence of such a relationship is called serial correlation and implies that either an important independent variable has been omitted or the wrong functional form has been used in the regression equation. Thus, rather than the regression equation explaining the underlying pattern of the data and having random error terms, these residuals still contain part of the basic pattern. The presence of serial correlation may be detected by noting a pattern in the residuals and by observing the value of the Durbin-Watson (D-W) statistic, calculated from the residuals themselves. Generally, for a data set of the size used in this study, a D-W statistic value between 1.5 and 2.5 implies a lack of serial correlation. The final basic assumption is that the residuals are normally distributed. If this assumption is not met, tests of significance and confidence intervals developed under the normality assumption will not be meaningful.

In arriving at a final choice for the set of independent variables to be included in the regression analysis, it was necessary to determine the appropriate lag for these variables. For example, an annual military pay raise in October may not influence enlistments until five or six months later. Similarly, a recruiter takes time to become proficient so

that an increase in canvassers one month may not have an impact until several months have passed. All of the independent variables previously explained (except DJ) were lagged from one to six months and several regressions were run to discover those variables that significantly contributed to the pattern of enlistments. The seemingly endless number of combinations was pared considerably by allowing a variable to enter the equation only once. For example, the civilian to military pay ratio (CMPR) could not be in the same equation at two different lags. The final selection was made based on the intuitive appeal of the regression equation and the statistical significance of the results. Using monthly enlistments over the period from December 1973 to June 1978 (Figure 3), a total of fifty-five observations, the multiple regression results were:

$$LSV = LN\left(\frac{S}{1-S}\right) = -2.669 - \underset{(11.655)}{5.647(CMPR5)} + \underset{(6.163)}{.100(UNEM)}$$

$$+ \underset{(3.33)}{.672(DJ)} + a \qquad (3.C.1)$$

where

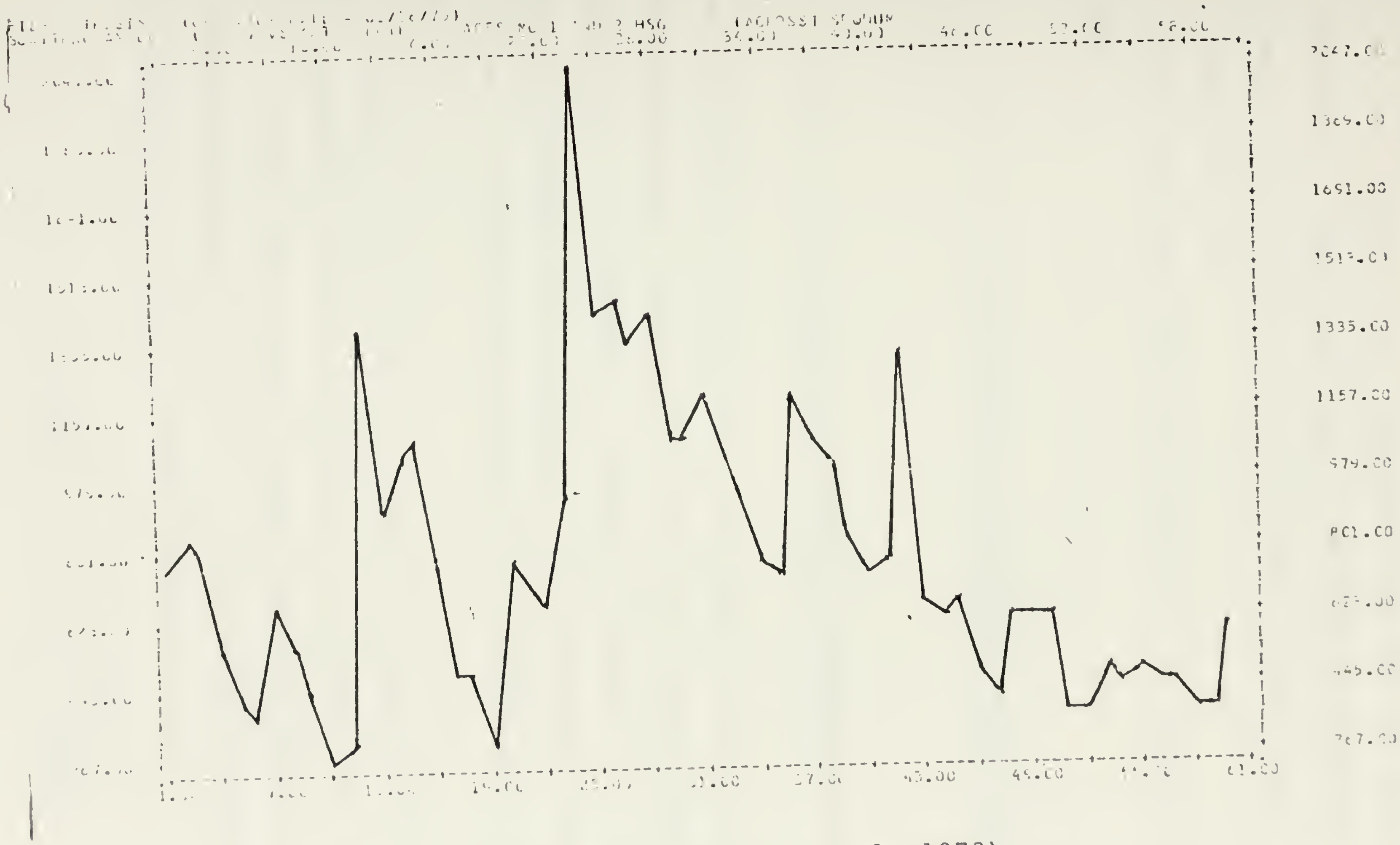
CMPR5 = civilian military pay ratio lagged 5 months

UNEM = unemployment with no lag

DJ = dummy variable for December of 1976

a = random error

MONTHLY ENLISTMENTS: JULY 1973 TO JUNE 1978



Case Number (e.g., 1 = July 1973)

Figure 3

The partial F-statistic appears in parenthesis below each coefficient. They are all significant at the 10% level. The percentage of the variation in monthly enlistments explained by the model, the coefficient of determination or R^2 value, was .42.

Writing the logit regression model in its nonlinear form one has:

$$S = \frac{1}{1 + e^{-[(-2.669-5.647(CMPR5)+.1(UNEM)+.672(DJ)+a)]}} \qquad (3.C.2)$$

$$S = \frac{1}{1 + e^{[2.669+5.647(CMPR5)-.1(UNEM)-.672(DJ)+a]}}$$

From the above it is clear that the signs of the coefficients of the explanatory variables are what one would expect intuitively. An increase in civilian relative to military pay will decrease the enlistment rate. Conversely, an increase in teenage unemployment will increase the enlistment rate.

A measure of the relative impact of changes in the explanatory variables on the enlistment rate is elasticity, the ratio of percentage changes between the enlistment rate and a given explanatory variable, X_i, with the others held constant. The formula for elasticity using the logit model is:

$$\varepsilon_i = B_i(1-S)X_i = B_iX_i \frac{e^{-\Sigma B_iX_i}}{1 + e^{-\Sigma B_iX_i}}$$

where

ε_i = elasticity of the enlistment rate with respect to explanatory variable X_i

B_i = regression coefficient of variable X_i

S = enlistment rate

X_i = explanatory variable i

Using June 1978 as a basis for calculating the elasticities with respect to pay and unemployment the following values were obtained:

elasticity with respect to pay	elasticity with respect to unemployment
-5.93	1.42

The results imply that if the civilian to military pay ratio was increased 1% and all other variables were held constant, the enlistment rate of male high school graduates in the top two mental categories would fall off by almost 6%. If teen-age unemployment increases 1%, the enlistment rate would increase about $1\frac{1}{2}$%. These elasticities are compared to those calculated from other studies shown in Table IV. Although the elasticity with respect to pay appears high compared with the results of these other studies, one should note that only in a log-linear model are the elasticities constant. Otherwise they depend on the regression coefficient and the values of the explanatory variables in the period for which the elasticity is calculated. Thus, it is difficult to compare

TABLE IV

STUDY	FUNCTIONAL FORM OF SUPPLY EQUATION	ELASTICITY WITH RESPECT TO MILITARY TO CIVILIAN PAY RATIO	ELASTICITY WITH RESPECT TO UNEMPLOYMENT RATE
Fisher	Semi-log	.62	.78
Klotz	Semi-log	1.47	3.53
Kim, et al.	Semi-log	2.78	3.18
Fechter	Linear	1.74	-1.40
Cook	Log-linear	2.23	1.36
Cooper	Log-linear	.97	.29
Cooper	Logit	1.25	.20
Amey, et al.	Linear	.784	1.27
Amey, et al.	Semi-log	.592	1.10

elasticities derived from various studies other than to get a sense of the impact of changes in the various explanatory variables on the enlistment rate.

The most surprising result of the regression was the exclusion of the variable representing recruiters, EFFREC. When forced into the equation, its regression coefficient was never statistically different from zero and had a negative sign indicating that the more recruiters in the field the less the enlistments. Intuition and experience would indicate that recruiters play a pivotal role in the enlistment decision of a quality applicant. However, it may be that such highly qualified applicants are shoppers and seek out the various services' recruiting offices themselves, thereby eliminating the "canvassing" part of recruiting which requires the manpower. Furthermore, the other services could be out-recruiting the Marine Corps in the quality market and adding additional recruiters has not changed this. In contrast, a loss of recruiters hasn't hurt either. Perhaps the mystique of the Marines coupled with the opportunity to serve in technical fields do not need to be extolled to the discriminating applicant.

The exclusion of the variable LEADS is not as disturbing. When forced into the equation its regression coefficient evidenced the same symptoms as EFFREC's. However, there was no way to disaggregate the variable in order to see how many of the post cards were returned by male high school

graduates in the top two mental categories. Consequently, there may indeed be no connection between these quality enlistments and the mass of returned postcards received each month.

The four basic assumptions of the multiple regression model were considered satisfied with the exception of the independence of the residuals. The Durbin-Watson statistic for the multiple regression model was 1.08, indicating positive serial correlation among the residuals. This is not surprising considering that no attempt to explain any seasonal pattern in the monthly enlistment rate was made. Combining the methodologies of multiple regression and Box-Jenkins is an excellent technique to correct this problem. The Box-Jenkins method is developed in the next section.

IV. THE BOX-JENKINS METHODOLOGY

A. THE ADVANTAGES OF BOX-JENKINS

The multiple regression analysis previously discussed required that a model be chosen on the basis of the analyst's experience or some pertinent economic interpretations. Use of the Box-Jenkins methodology does not require that such a specific model be chosen. Instead, the researcher applies a specific technique to analyze the data and iteratively eliminate inappropriate models until he is left with the most suitable one. The Box-Jenkins methodology is particularly well suited to analyzing a complex time series in which a number of observations, influenced by seasonal and/or cyclical factors, are taken over discrete, equally spaced periods of time, and a forecast of some future time period is desired. In an attempt to discern the underlying pattern in the time series, the general method illustrated in Figure 4 is applied [Ref. 13]. It should be noted that the Box-Jenkins forecasting method shares with other time series techniques the fundamental assumption that these past patterns of the data will repeat themselves in the future.

B. THE GENERAL APPROACH

The Box-Jenkins forecasting method is composed of four stages. In stage 1 a specific model that can be tentatively entertained as the one best describing the situation under analysis is chosen. Stage 2 involves fitting that model to

THE BOX-JENKINS FORECASTING METHOD

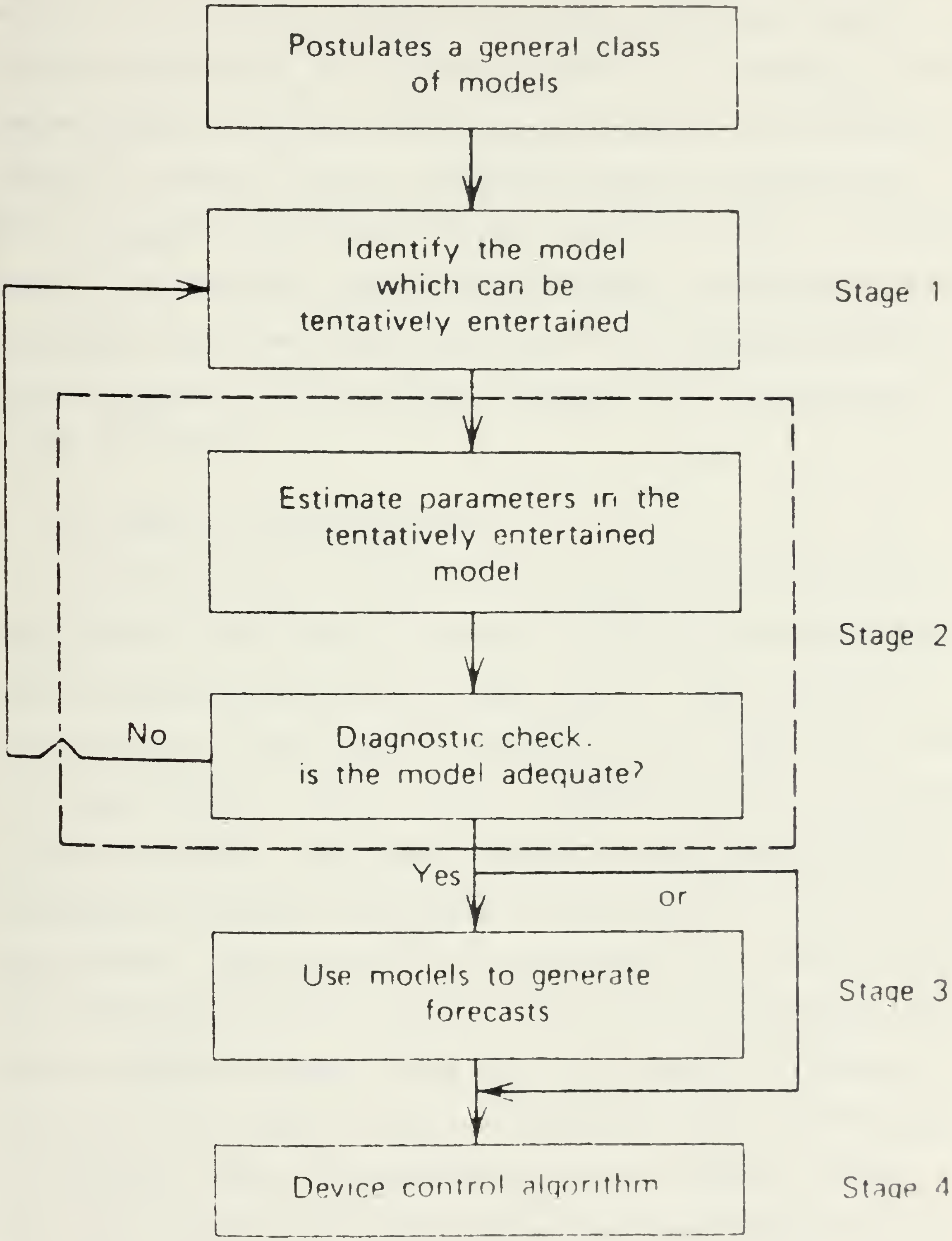


Figure 4

the available historical data (in this case regular Marine Corps monthly enlistments) and analyzing the deviations between the observed values of the time series and those values predicted by the tentative model. If the model is not deemed adequate by this statistical procedure, the approach returns to stage 1 and an alternative model is specified. Once an appropriate model has been isolated, stage 3 and then stage 4 are pursued. These latter stages involve developing a forecast for some future time period and implementing a plan for updating the model and acting on the implications of the forecasts.

C. THE ROLE OF AUTOCORRELATION

In constructing a model of a time series one needs to describe the relationship between a current observation and previous observations of the same series. This concept of correlation is a key tool in arriving at the underlying pattern of a time series. Correlation is a measure of the association of two variables. When these variables are observations from the same time series the degree of association is called autocorrelation. Autocorrelation is measured by an autocorrelation coefficient which like the correlation coefficient, may take on values between -1 and +1. For example, if observations in a time series which are above the mean are followed by similarly valued observations four time periods later, one would say that the observations four periods apart are positively autocorrelated. The closer the autocorrelation

coefficient is to +1 the stronger the positive autocorrelation. That is, if the last observation in a time series was above the series mean, the strong positive autocorrelation at four periods in our hypothetical example would suggest that a practical choice for a forecast four periods in the future would also be a value above the series mean. However, autocorrelation does not imply causality.

Thus, autocorrelations provide important information about the pattern underlying a time series. In a set of completely random observations the autocorrelation among successive values will theoretically be zero, whereas observations of strong seasonal or cyclical character will be highly autocorrelated. Figure 5 presents the autocorrelations, denoted r_k, of different time lags of monthly gasoline consumption in Great Britain [Ref. 13]. These autocorrelations reveal a strong seasonal pattern of 12 months duration since the highest results among the set of autocorrelation values occur every 12 successive months. It is this kind of information, derived from the plot of the autocorrelations for various lags (called a correlogram), that can be utilized by the Box-Jenkins approach to arrive at the optimal forecasting model. Note that no assumptions need be made about the observations or their pattern to calculate the autocorrelation coefficients. The correlogram may be used to reveal the type of time series under analysis and its pattern.

CORRELOGRAM FOR GASOLINE SALES IN GREAT BRITAIN

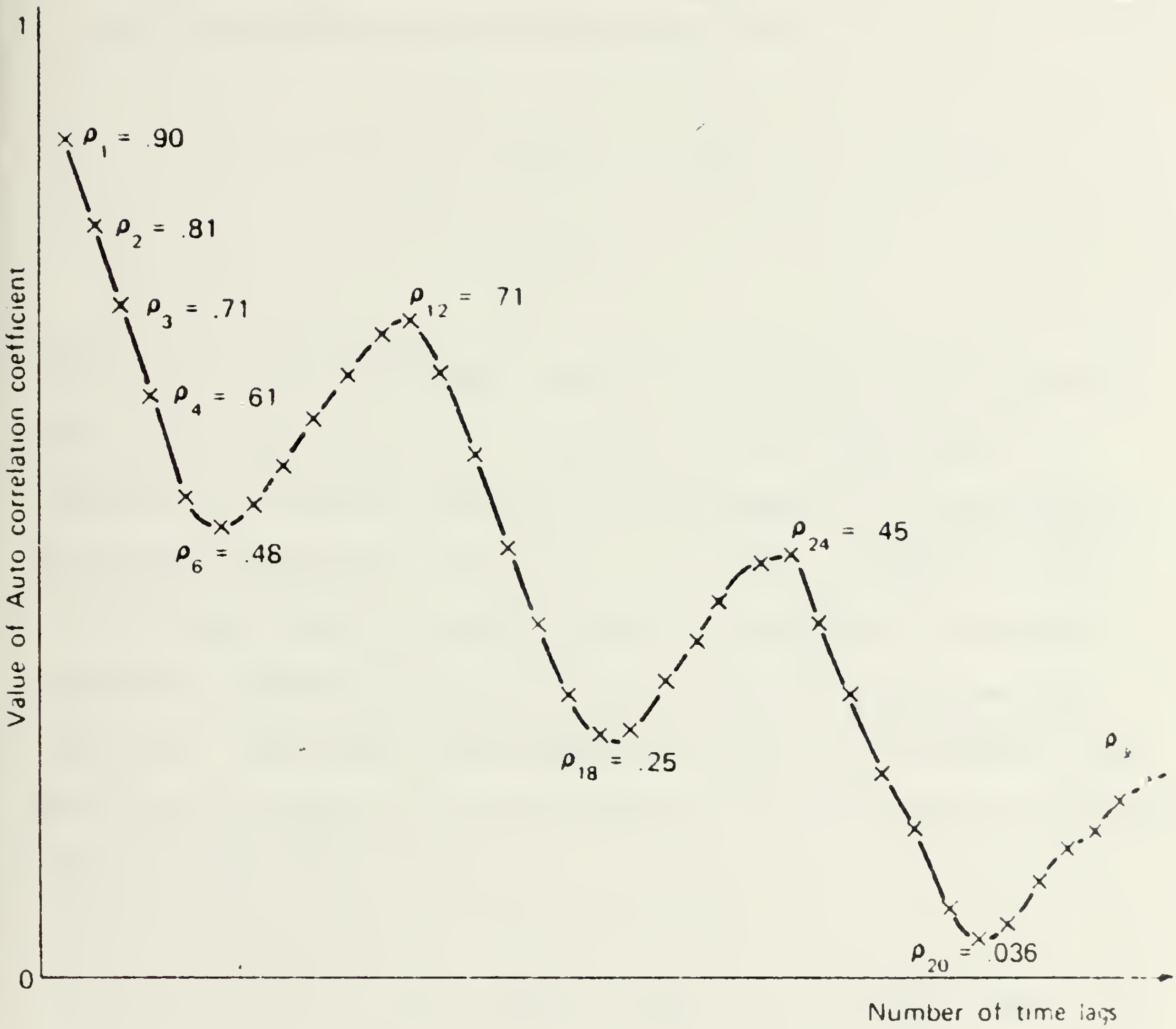


Figure 5

D. THE MODEL TYPES

The Box-Jenkins method is based on three general classes of models for any time series: (1) the autoregressive or AR model, (2) the moving average or MA model, and (3) the mixed autoregressive-moving average or ARMA model.

An autoregressive model has the form:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \cdots + \phi_p Y_{t-p} + \theta_o + e_t$$

where Y_t is the dependent variable, such as monthly enlistments, and $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$ are the independent or explanatory variables which in this example are enlistments from previous months (t-1, t-2, ..., t-p). Finally, there is a constant term θ_o and an error or residual term which represents random deviations that cannot be explained by the model. The term autoregressive is used to describe the above model since it is very similar to the multiple regression equation described earlier:

$$Y = B_o + B_1X_1 + B_2X_2 + \ldots + B_pX_p + e_t$$

Now, however, the explanatory variables are replaced by lagged values from the time series itself. The model postulates that the enlistments in month t are influenced by those of the past p months. The strength of a past month's influence

is weighted according to its coefficient ϕ_i where i ranges from 1 to p.

Unfortunately, not all time series may be handled with just the pure autoregressive model. Some data are more properly described using the moving average or MA model:

$$Y_t = \theta_o + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

where Y_t is again a time series observation such as monthly enlistments and θ_o is a constant term. The model asserts that the dependent variable is a function of the current and previous values of the error term. The negative signs preceding the MA coefficients are merely a convention.

The final general class of time series models is the mixed autoregressive moving average model or ARMA:

$$Y_t = \theta_o + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

This model presumes that future monthly enlistments, for example, depend both on historical enlistments and the past deviations or errors between those enlistments forecasted by the model and those actually observed.

In order to simplify the notation involving the various time series models, the backshift operator is introduced. The backshift operator is defined as B^k where the exponent k

determines the amount or degree of the lag. For example, $BY_t = Y_{t-1}$ and $B^3Y_t = Y_{t-3}$. In general, $B^kY_t = Y_{t-k}$. The backshift operator may be combined in a polynomial such as

$$(1-B-B^2)Y_t = Y_t - Y_{t-1} - Y_{t-2} .$$

Furthermore, standard mathematical operations apply to the backshift operator. For example, $B(1-B)Y_t = (B-B^2)Y_t = Y_{t-1} - Y_{t-2}$. The three general models using the backshift notation are:

(1) autoregressive of order p or AR(p)

$$(1-\phi_1B - \phi_2B^2 - \ldots - \phi_pB^p)Y_t = \theta_o + e_t$$

(2) moving average of order q or MA(q)

$$Y_t = \theta_o + (1 - \theta_1B - \theta_2B^2 - \ldots - \theta_qB^q)e_t$$

(3) mixed autoregressive moving average of order p and q or ARMA(p,q)

$$(1 - \phi_1B - \phi_2B^2 - \ldots - \phi_pB^p)Y_t = \theta_o + (1 - \theta_1B - \theta_2B^2 - \ldots - \theta_qB^q)e_t$$

The choice of one of the three models and its degree can be made by examining the autocorrelations and a set of related statistics, the partial autocorrelations. The procedure

involves selecting a tentative model based on an examination of the correlogram, calculating initial estimates of the coefficients or parameters (the ϕ's, θ's, and θ_o), and finally examining the autocorrelations of the models' residuals. If the model has been correctly fitted to the time series data then these differences between the actual and predicted values should be randomly distributed over the time series. That is, their autocorrelations should be close to zero and exhibit no pattern. However, should a pattern be present the Box-Jenkins methodology provides for developing a new model and repeating the analysis until the residuals are randomly distributed. Then the final model may be used for forecasting and policy decisions.

V. THE APPLICATION OF BOX-JENKINS TO MONTHLY MARINE ENLISTMENTS

A. THE MODEL, STAGE 1

The Box-Jenkins procedure was applied to a time series of monthly male enlistments in the regular Marine Corps over the period from July of 1973 to June of 1978, a total of sixty observations. In order to complement the multiple regression analysis, the same data were used: the monthly enlistments of high school graduates in the top two mental categories as determined by the applicants' scores on the Armed Services Vocational Aptitude Battery of tests (ASVAB). The monthly enlistments, variable VSMART, are listed in Table III and shown graphically in Figure 3.

The first step was to calculate the autocorrelations and partial autocorrelations for the original data (Figs. 6 and 7). The exponentially decaying sine wave pattern indicates that the time series may be described by a mixed autoregressive moving average model, i.e., ARMA(p,q). Fortunately, most time series may be adequately described by models of low order. That is, p and q generally take on values of less than three. Prior to calculating any initial estimates of the parameters for the ARMA(p,q) model a seasonal difference of twelve months was imposed on the time series. From the plot of the monthly enlistments in Figure 3, such a seasonal pattern is evident. This differencing operation results in a "smoother" time series which may be more accurately

AUTOCORRELATIONS FOR MONTHLY ENLISTMENTS

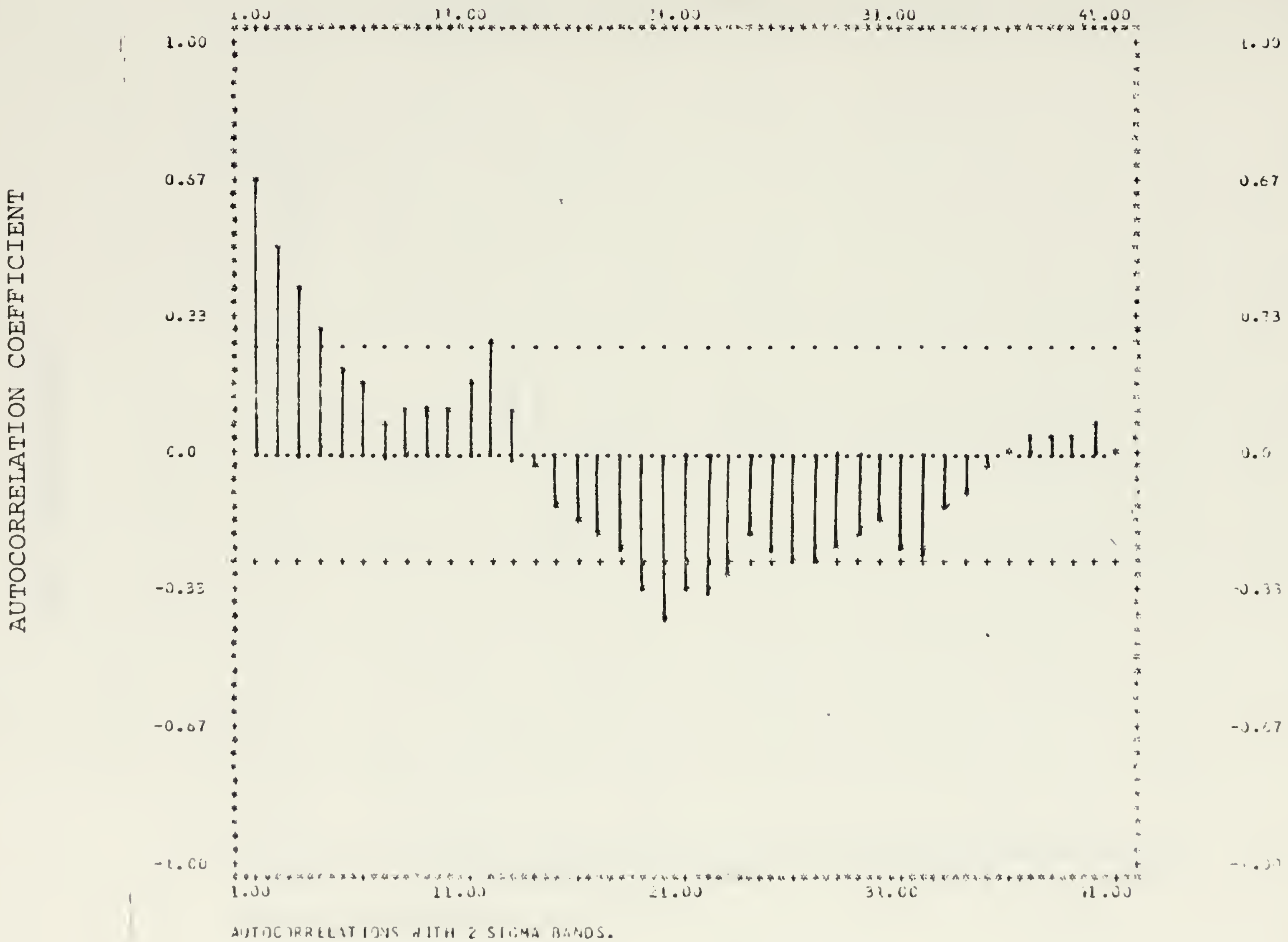


Figure 6

PARTIAL AUTOCORRELATIONS FOR MONTHLY ENLISTMENTS

PAUTO COEFFICIENTS

LAG

Figure 7

represented by the model ultimately chosen. Using the backshift operator notation, the time series has been transformed to:

$$(1 - B^{12})Y_t = Y_t - Y_{t-12}$$

B. THE MODEL, STAGE 2

Having selected a tentative ARMA(p,q) model, the parameters (ϕ's, θ's, θ_o) were estimated and the autocorrelation function for the residuals was analyzed. After several iterations of this procedure it was apparent that the following model adequately described the time series of monthly enlistments:

$$(1 - \phi_1 B)(1 - B^{12})Y_t = \theta_o + (1 - \theta_1{}^s B^{12})e_t$$

where Y_t is the number of male high school graduates in mental categories I and II who enlisted in month t. The constant is denoted by θ_o and the error term by e_t. The autoregressive coefficient is ϕ_1 and the moving average coefficient is $\theta_1{}^s$ where the exponent s merely indicates that the MA coefficient is associated with a backshift operator of seasonal lag twelve.

Multiplying the backshift operator polynomials together the expression becomes:

$$(1 - \phi_1 B - B^{12} + \phi_1 B^{13})Y_t = \theta_o + (1 - \theta_1{}^s B^{12})e_t$$

Rearranging terms and carrying through the backshift operator the model is:

$$Y_t - \phi_1 Y_{t-1} - Y_{t-12} + \phi_1 Y_{t-13} = \theta_o + e_t - \theta_1{}^s e_{t-12}$$

$$Y_t = \phi_1 Y_{t-1} - \phi_1 Y_{t-13} + Y_{t-12} + \theta_o + e_t - \theta_1{}^s e_{t-12}$$

The interactive computer program entitled the Time Series Editor [Ref. 14] was used to perform the calculations and plot the various results. The final Box-Jenkins model for the enlisted time series was:

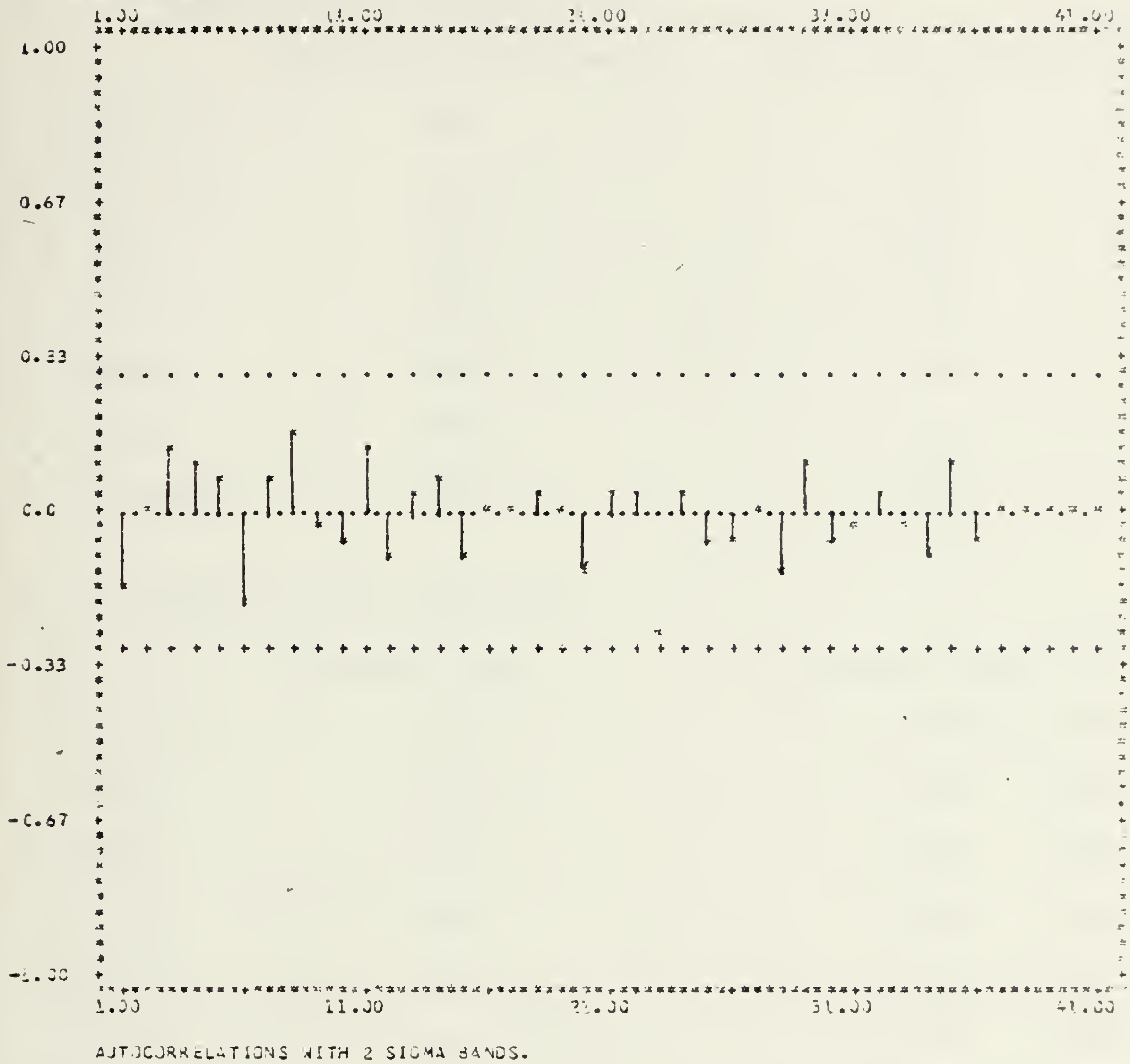
$$Y_t = .804251Y_{t-1} + Y_{t-12} - .804251Y_{t-13} - 7.095902 + e_t - .478727e_{t-12} \quad (4.B.1)$$

The coefficients were statistically significant at the 5% level. The correlogram for the residuals of the above model is plotted in Fig. 8. The autocorrelations appear to be randomly distributed along the time series and their values are all near zero.

C. FORECASTING, STAGE 3

Having determined an adequate model by noting that the autocorrelations between the residuals had no discernible pattern and their values were all near zero, forecasts of enlistments were made. Table V and Figure 9 show the results.

CORRELOGRAM FOR RESIDUALS FROM MODEL OF ENLISTMENT DATA



LAG

$$(1-\phi_1 B)(1-B^{12})Y_t = \theta_o + (1-\theta_1{}^s B^{12})e_t$$

Figure 8

TABLE V

BOX-JENKINS TIME SERIES FORECASTS

A. Forecast Origin December 1977

	Observed enlistments	Forecast (Jan-Jun 1978)
JAN 78	487	252
FEB 78	448	315
MAR	455	365
APR	362	248
MAY	384	289
JUN	590	811
Average	454	380

B. Forecast Origin June 1978

	Previous year	Forecast (July 1978 - June 1979)	
JUL 77	614	JUL 78	497
AUG	615	AUG	550
SEP	390	SEP	426
OCT	364	OCT	393
NOV	470	NOV	429
DEC	436	DEC	562
JAN 78	487	JAN 79	469
FEB	448	FEB	453
MAR	455	MAR	460
APR	362	APR	339
MAY	384	MAY	356
JUN	590	JUN	704
Average	468	Average	470

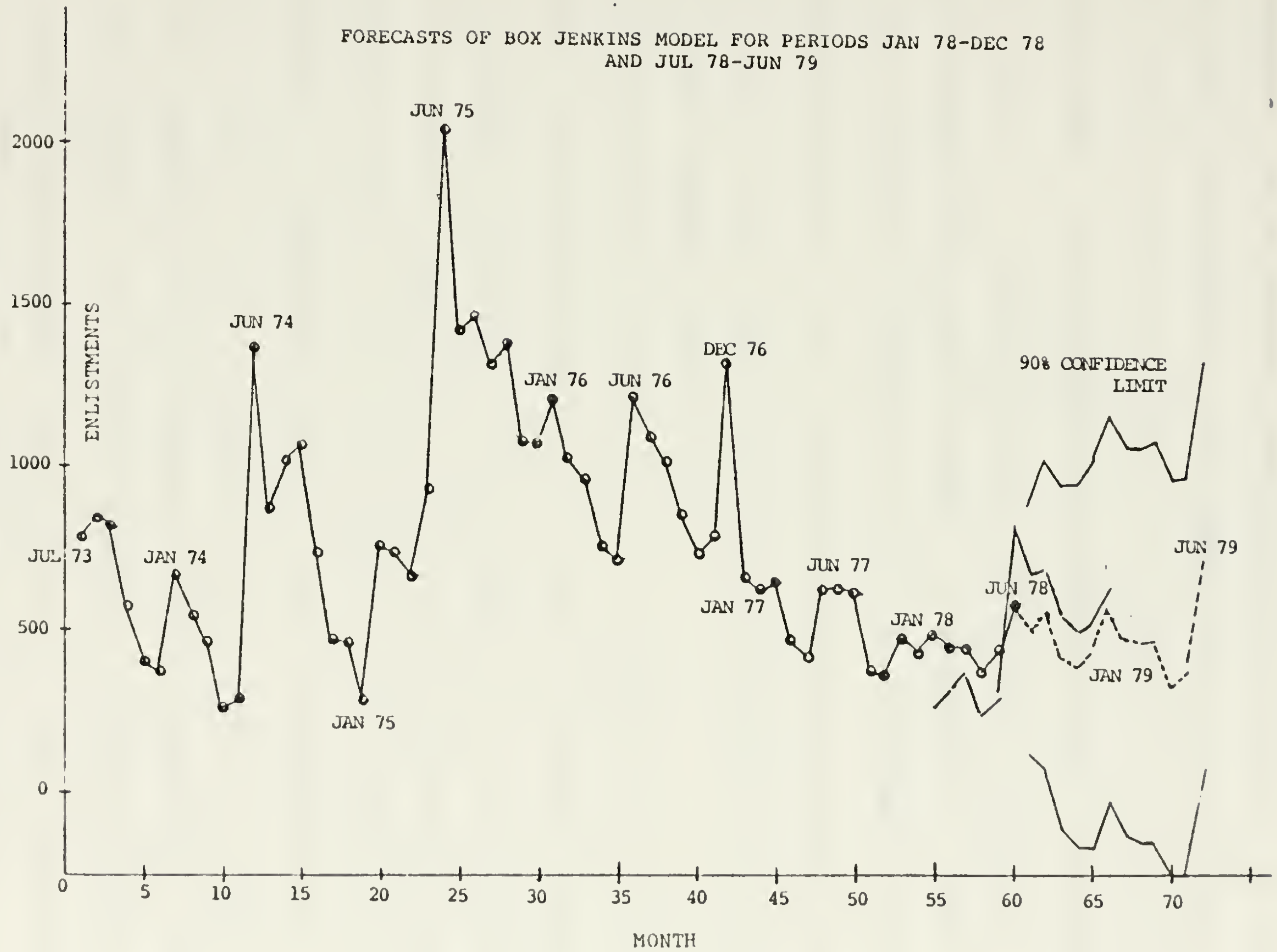

FORECASTS OF BOX JENKINS MODEL FOR PERIODS JAN 78-DEC 78
AND JUL 78-JUN 79
ENLISTMENTS
2000
1500
1000
500
0
JUL 73
JAN 74
JUN 74
JAN 75
JUN 75
JAN 76
JUN 76
DEC 76
JAN 77
JUN 77
JAN 78
JUN 78
JAN 79
JUN 79
90% CONFIDENCE
LIMIT
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
MONTH

Figure 9

In part A of the table the forecast origin was chosen as December of 1977 in order to have an overlap of six months between the actual number of enlistments and those forecasted for a twelve month period. Although a month by month comparison between the known and forecasted enlistments is not particularly enlightening, the model does capture the pattern of enlistments for that period. It should also be noted that the model incorporates enlistments lagged thirteen months which, in forecasting the number for January of 1978, would encompass the large "spike" at December of 1976 (Figure 9). This anomaly, caused by a rush of enlistments to take advantage of the G.I. Bill benefits before they expired on 31 December of that year, certainly distorted the forecasts for January through June of 1978.

The forecasting results using a forecast origin of June 1978 are much more encouraging. Part B of Table V compares the enlistment forecasts for the period July 1978 through June 1979 with an identical period of the previous year. It is evident from the table and Figure 9 that the pattern of enlistments is repeated quite well by the model given in equation 4.B.1. The confidence limits drawn on Figure 9 indicate that one may be 90% certain that the predicted value will fall within the indicated bounds. These statistical calculations confirm the intuitive notion that the farther out in time a forecast is made the more susceptible to error it becomes.

VI. A COMBINED MODEL AND FORECAST COMPARISON

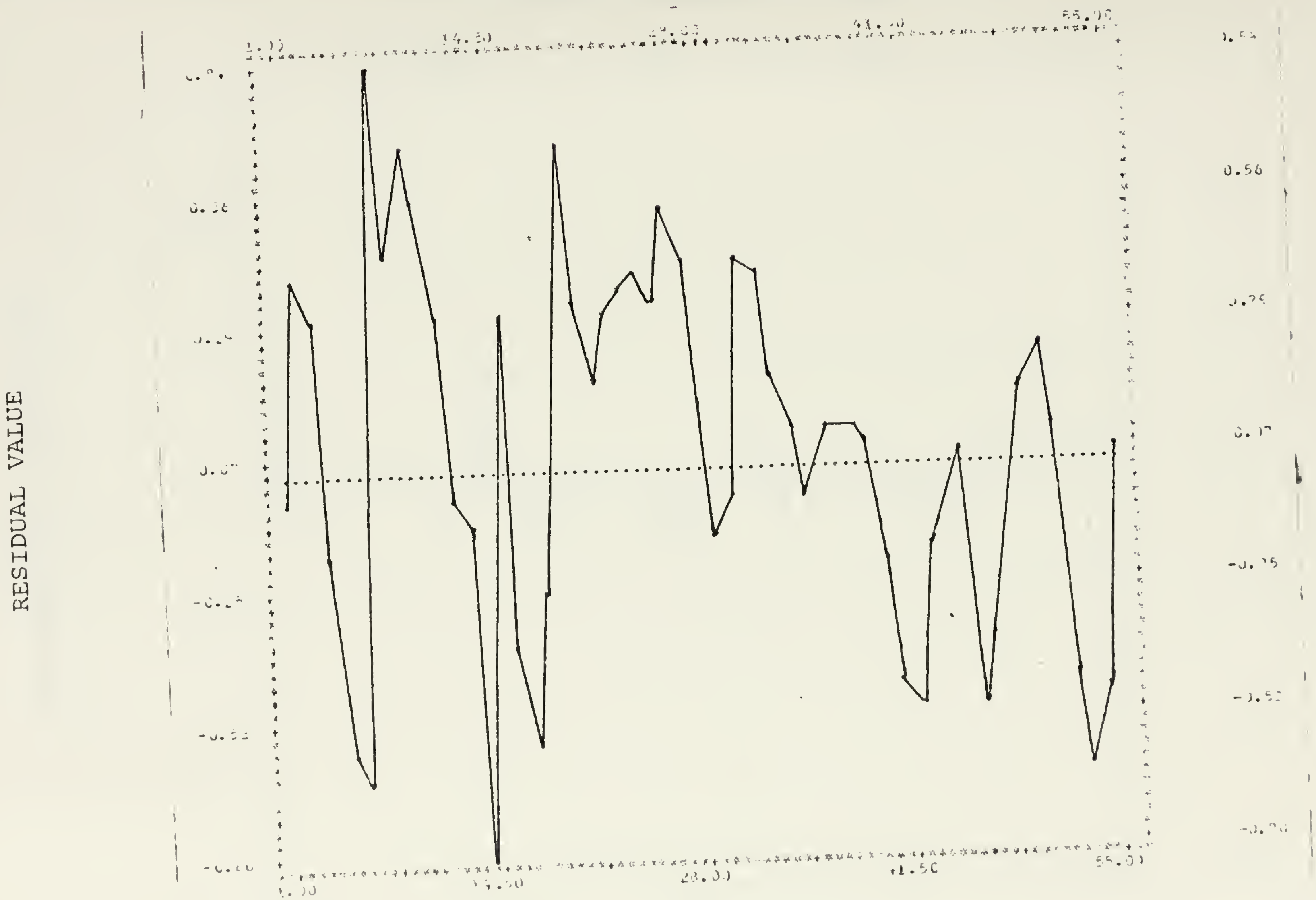
A. DEVELOPING THE COMBINED MODEL

The multiple regression model developed in Section III satisfied all the basic assumptions except for the independence of the residuals. In order to correct this last problem, the Box-Jenkins methodology was applied to the residuals as if they were an original time series. Once an appropriate model for the residuals was derived, it was combined with the multiple regression model to forecast enlistments more accurately.

The residuals from the multiple regression model (equation 3.C.1) were plotted along with their autocorrelations and partial autocorrelations (Figs. 10, 11, and 12). From an analysis of the autocorrelations a seasonal difference of 12 months was taken. Then the correlogram for the differenced residuals was plotted (Figs. 13 and 14). The decaying sine wave pattern and the first two large spikes in the partial autocorrelations suggested an AR(2) model. The final plot of the autocorrelation function for the error terms from this model indicates they are randomly distributed and near zero (Fig. 15). The model for the residuals was:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B^{12}) z_t = \theta_o + e_t$$

RESIDUALS FROM MULTIPLE REGRESSION MODEL



Case Number (e.g., 1 = December 1973)

Figure 10

CORRELOGRAM OF RESIDUALS FROM
MULTIPLE REGRESSION MODEL

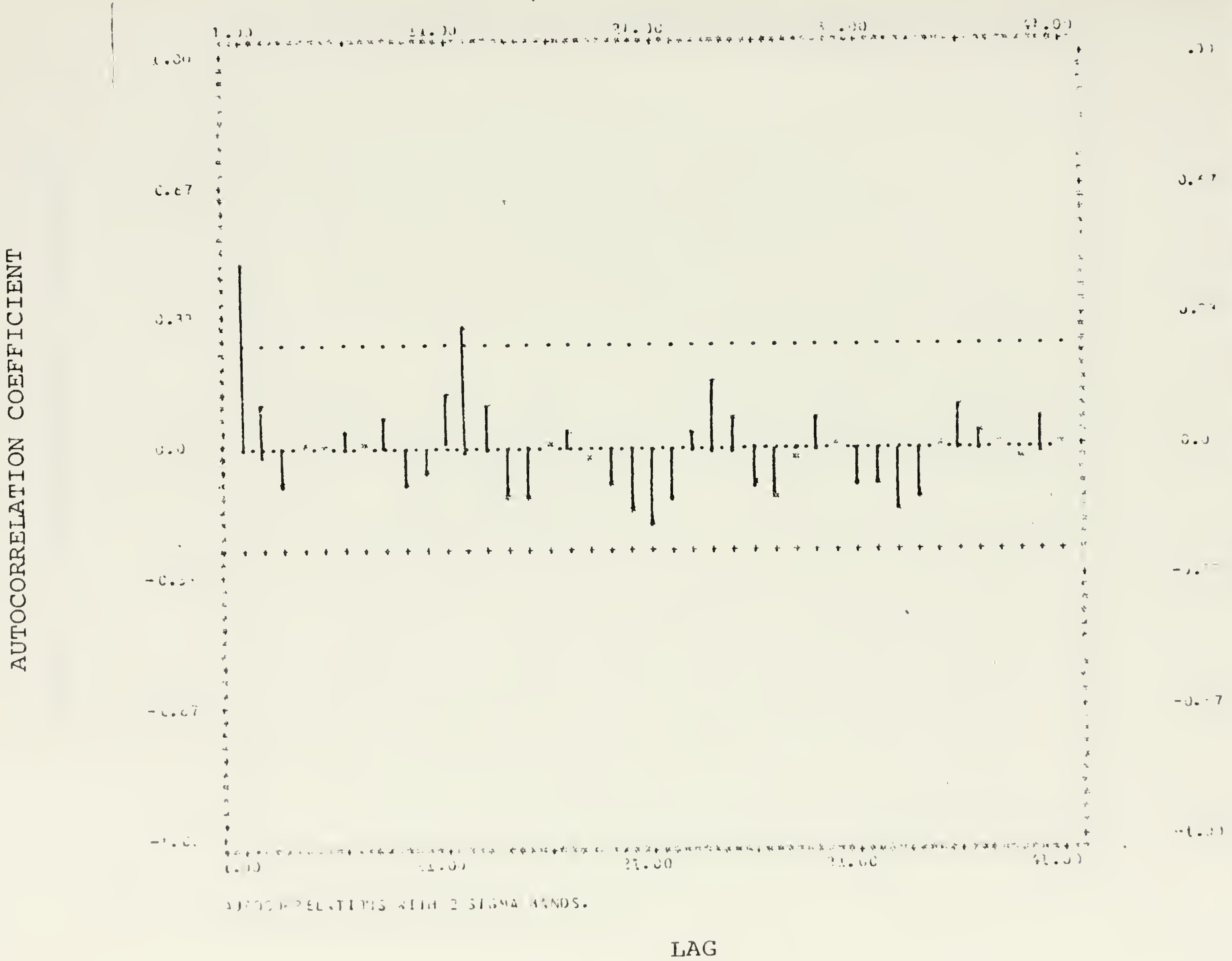


Figure 11

PARTIAL AUTOCORRELATIONS OF RESIDUALS FROM
MULTIPLE REGRESSION MODEL

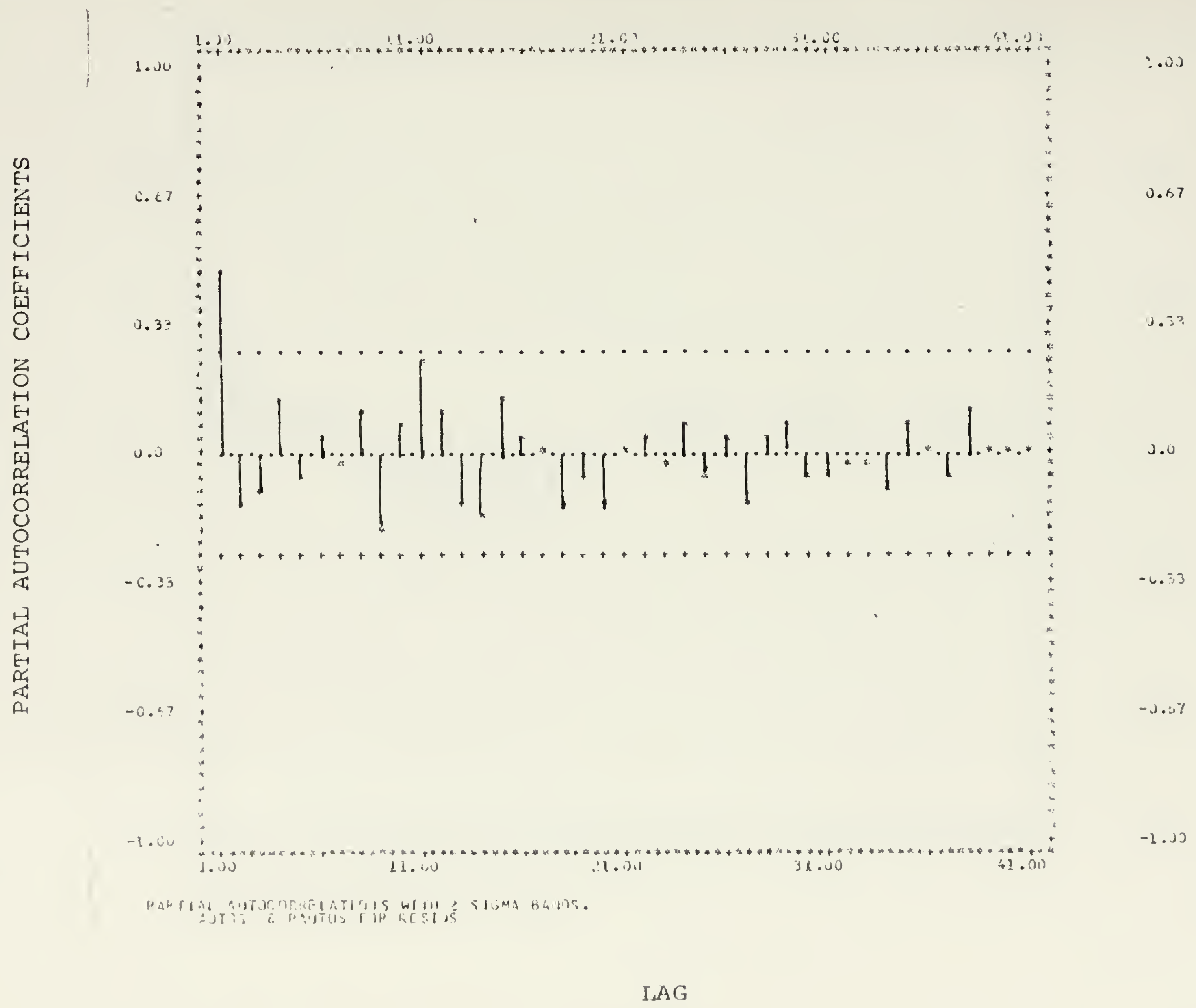


Figure 12

CORRELOGRAM FOR DIFFERENCED RESIDUALS

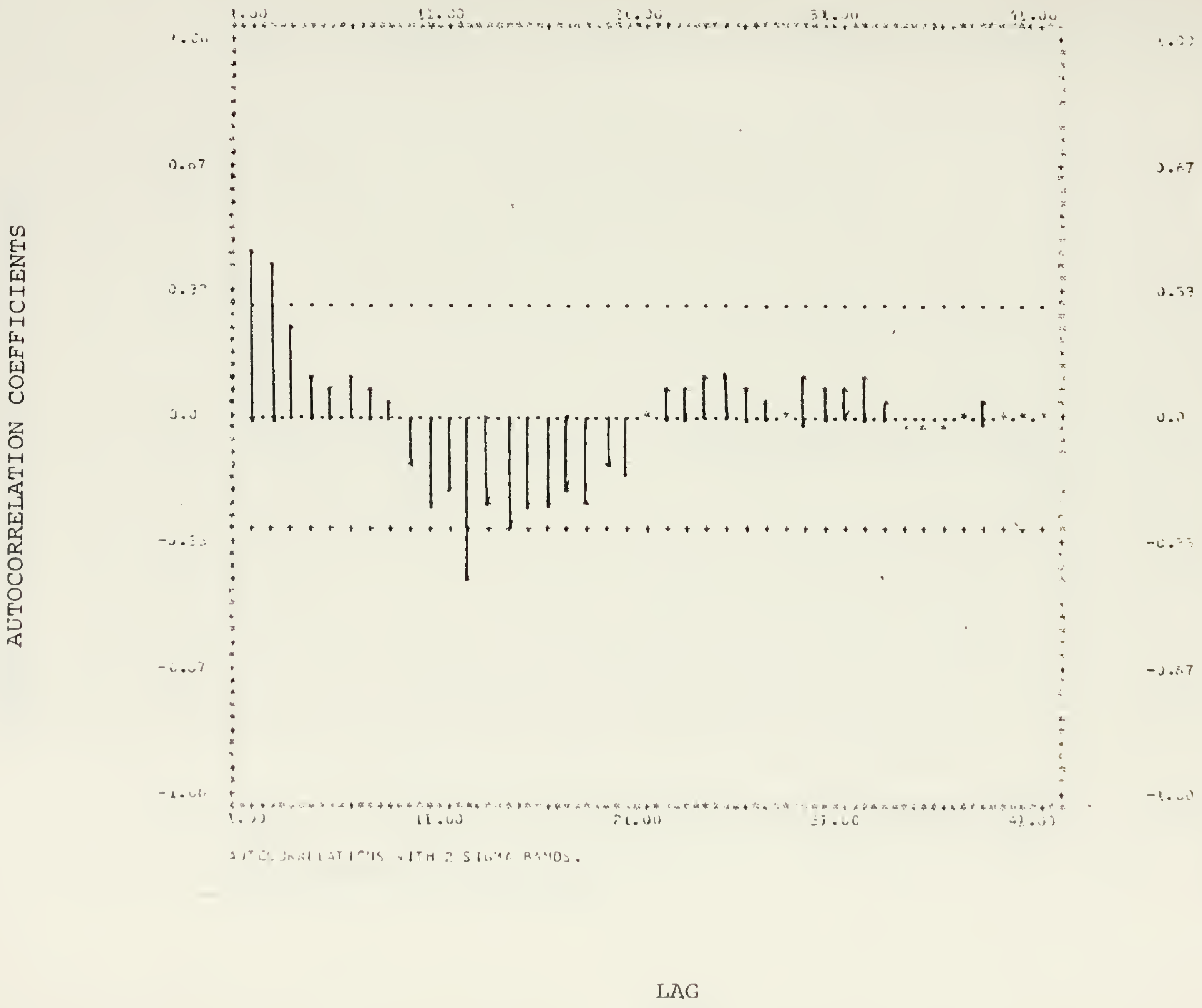


Figure 13

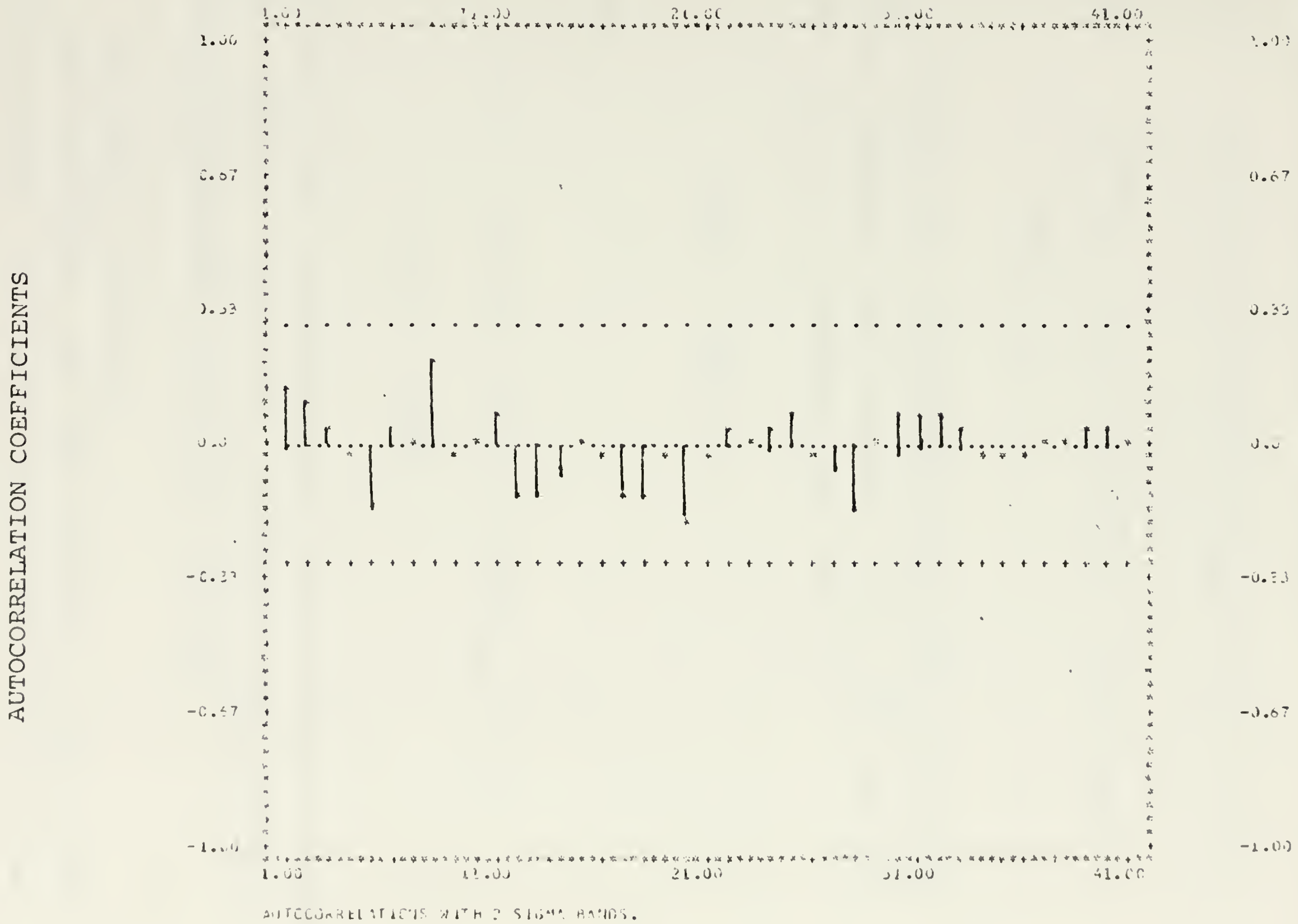
PARTIAL AUTOCORRELATIONS FOR DIFFERENCED RESIDUALS

PARTIAL AUTOCORRELATION COEFFICIENTS

LAG

Figure 14

CORRELOGRAM FOR ERROR TERMS FROM
MODEL OF RESIDUALS



LAG
Figure 15

where Z_t is the residual value from the multiple regression model and e_t is the error term for the autoregressive model. Expanding and rearranging terms, one has:

$$Z_t = \phi_1(Z_{t-1} - Z_{t-13}) + \phi_2(Z_{t-2} - Z_{t-14}) + Z_{t-12} + \theta_o + e_t$$

$$Z_t = .3407(Z_{t-1} - Z_{t-13}) + .2596(Z_{t-2} - Z_{t-14}) + Z_{t-12}$$
$$- .03683 + e_t$$

Let LSV_{MR} equal the predicted values using the multiple regression equation alone (equation 3.C.1). Then the predicted value for the combined model, LSV_c, is:

$$LSV_c = LSV_{MR} + Z_t'$$

where

$$Z_t' = .3407(Z_{t-1} - Z_{t-13}) + .2596(Z_{t-2} - Z_{t-14})$$
$$+ Z_{t-12} - .03683$$

and the Z_i's are residuals derived from the multiple regression model. That is, they are the differences between the actual value of the dependent variable LSV and LSV_{MR}.

B. THE FORECAST COMPARISONS

The various models have been developed in order to forecast enlistments. The period January 1978 to June 1978 was

chosen as a forecast horizon since the actual enlistments for these months were known. Table VI shows the results. As measured by the root mean squared error, the combined model is the best by far. However, the Box-Jenkins model does include enlistments lagged thirteen months which for this forecast period involves the non-recurring surge of enlistments in December, 1976. This certainly influenced the forecasted values (see Table V). The multiple regression equation still contains serial correlation in the residuals, indicating that the pattern of the time series had not been fully accounted for by such a model. Therefore, it was not unexpected that the combined model more accurately forecasts enlistments over the period chosen.

Unfortunately, the difficulty in forecasting with a model involving multiple regression is one must also forecast the values of the explanatory variables such as unemployment and the civilian to military pay ratio. Thus, there is a compounding negative effect on the accuracy of such forecasts. The attractiveness of the Box-Jenkins approach is that the analyst may let the time series speak for itself. However, the basis of time series analysis is that the underlying pattern will repeat itself in the future. So if the time series of monthly enlistments contains such non-recurring anomalies as the surge in December, 1976, the forecasting results will be degraded.

TABLE VI

FORECAST COMPARISONS
(residuals in parentheses)

Date	Actual Enlistments	Predicted Value Mult. Regression Plus Box-Jenkins	Predicted Value Box-Jenkins	Predicted Value Mult. Regression
Jan 78	487	413 (74)	252 (235)	350 (137)
Feb 78	448	466 (-18)	315 (133)	332 (116)
Mar 78	455	589 (-134)	365 (90)	450 (5)
Apr 78	362	391 (-29)	247 (115)	615 (-253)
May 78	384	321 (63)	289 (95)	625 (-241)
Jun 78	590	441 (149)	811 (-221)	545 (45)
Root mean squared error of residuals		101	174	177

$$\sqrt{\frac{\Sigma(Y_t - \hat{Y}_t)^2}{n-1}}$$

Y_t = actual value

$\hat{Y}_t$ = value predicted by model

n = 6. (number of observations)

VII. CONCLUSION

A. SUMMARY

Three models using two distinct analytical techniques have been developed. A multiple regression model was derived based on its compatibility with a theory of occupational choice, the intuitive appeal of the explanatory variables, the past literature of manpower supply, and the statistical significance of each variables impact on monthly enlistments (equation 3.C.2). The Box-Jenkins methodology was applied to the time series of monthly enlistments and a model of the underlying pattern in the data was developed. As a further refinement the residuals from the multiple regression equation were treated as an original time series and the Box-Jenkins technique applied to them. Then the two models were combined and forecasts calculated.

A comparison of the enlistment forecasts made by each of the models indicates that the combined model is best as measured by the root mean square of the residuals. This is undoubtedly true for short term forecasts of say only three months. A greater forecast horizon favors the Box-Jenkins approach since it is not necessary to also forecast teenage unemployment or civilian pay to calculate the predicted enlistments. There is somewhat of a dilemma in that multiple regression allows the analyst to explore causal relationships but hampers the important job of forecasting while the

Box-Jenkins approach characterizes just the opposite. A compromise may be reached by using the combined model for short term forecasts and the Box-Jenkins approach for longer projections where guesses at future values of the explanatory variables are not necessary.

B. POLICY IMPLICATIONS OF THE RESULTS

The elasticities calculated from the multiple regression analysis indicate areas for policy emphasis. However, the choice of high school graduates aged 16 to 24 not enrolled in college as the population cohort by which the monthly enlistments were deflated upwardly biased the elasticities. This may be seen from the logit expression for elasticity:

$$\varepsilon_i = b_i X_i (1 - S)$$

where

ε_i = elasticity with respect to X_i

b_i = regression coefficient

X_i = explanatory variable

S = enlistment rate

Using such a large population makes the enlistment rate very close to zero. A more accurate deflater, had it been available, would have been the percentage of that population cohort in the top two mental categories. This smaller number would

have increased the enlistment rates thereby lowering the elasticities.

Nevertheless, the strong influence of the civilian to military pay ratio on the enlistment decision of applicants who are high school graduates and in the top two mental categories indicates that military pay plays a crucial role in attracting quality enlistees. This is a variable that while not directly under the control of the services may certainly be lobbied for. It may be that a salary system with "in kind" benefits converted to cash compensation would significantly increase the enlistees from this needed cohort of young men.

The significant impact of teenage unemployment should also be considered when assessing the Marine Corps' recruiting efforts. Mathematically, a high unemployment rate is good for the All Volunteer Force. An increasing minimum wage and employers' preferring older workers are helping to maintain this indirect benefit to recruiting. It is ironic that the services are undergoing recruiting shortfalls during periods of high teenage unemployment. This should indicate that if teenage unemployment is reduced to low levels as envisaged in the recently passed Humphrey-Hawkins Full Employment Act, then an even worse recruiting picture will be in the offing. The Marine Corps and the other services as well should promote cooperation with state and federal employment agencies. While some states prefer not to act

as recruiters for the federal government, the overriding concern should be to employ the teenage job seeker.

The absence of recruiters as an explanatory variable in the regression model is puzzling. It could be that the higher quality applicant, like one at the other end of the scale, needs no personalized persuasion to enlist. He may shop the services on his own and then make his occupational choice regardless of the size of thé recruiting force. Additionally, the other services' efforts in this market could overshadow those of the Marine Corps. This would imply that the monthly enlistments under study have come primarily from those already predisposed to the Marine Corps. That is, they would have enlisted by walking into a recruiting office somewhere; an act which is independent of the number of recruiters in the area.

C. IDEAS FOR FUTURE RESEARCH

Certainly the models should be updated and refined. The Box-Jenkins methodology optimally requires over 100 observations. Hence the time series should be extended as data becomes available and the analysis repeated. Furthermore, although this study was done on a national basis, if data for the explanatory variables is available, the same analysis techniques may be applied at the District or Recruiting Station levels. The forecasts produced from continually updated models could be used to project recruiting depot workloads, school inputs, skill shortages and overloads,

logistic needs, and allocation of recruiting resources. Additionally, the variables in the multiple regression model could be experimented with and other functional forms could be tried. Managerial forecasting using multiple regression analysis and the Box-Jenkins technique is a developed field of analysis which the Marine Corps may profitably put to use in many areas where the projection of current values is important.

BIBLIOGRAPHY

1. General Research Corporation Draft, Supply Estimation of Enlistees to the Military, by D.M. Amey, A.E. Fechter, D.W. Grissmer, and G.P. Sica, June 1976.

2. Cook, A.A., Jr., "Supply of Air Force Volunteers," The Presidents Commission on an All-Volunteer Armed Force, V. 1, p. II-4-2, November 1970.

3. Rand Corporation Report R-1450-ARPA, Military Manpower and the All-Volunteer Force, by Richard V.L. Cooper, September 1977.

4. Institute for Defense Analyses Report P-845, Army Enlistments and the All-Volunteer Force: The Application of an Econometric Model, by A.E. Fechter, February 1972.

5. Fechter, A.E., "Impact of Bay and Draft Policy on Army Enlistment Behavior," The President's Commission on an All-Volunteer Armed Force, V. 1, November 1970.

6. Gilman, H.J., "The Supply of Volunteers to the Military Services," The President's Commission on an All-Volunteer Armed Force, V. 1, p. II-1-10, November 1970.

7. Gray, B.C., "Supply of First-Term Military Enlistees: A Cross-Section Analysis," The President's Commission on an All-Volunteer Armed Force, V. 1, II-2-1, November 1970.

8. George Washington University Report 1142, "Compensation and Non-Compensation Inducements and the Supply of Military Manpower," by S.E. Haber, 30 July 1973.

9. Hause, J.C., "Enlistment Rates for Military Service and Unemployment," Journal of Human Resources, V. 8, Winter 1973.

10. Center for Naval Analyses Report RC-235, Navy Recruiting in an All-Volunteer Environment, by C. Jehn and H.E. Carroll, July 1973.

11. Lippman, S.A. and McCall, J.J., "The Economics of Job Search: A Survey," Economic Inquiry, V. 14, June and September 1976.

12. Lucas, R.E.B., "Hedonic Wage Equations and Psychic Wages in the Returns to Schooling," American Economic Review, V. LXVII, pp. 549-558, September 1977.

13. Makridakis, S. and Wheelwright, S.C., Forecasting Methods for Management, 2d ed., John Wiley & Sons, 1977.

14. Naval Postgraduate School Report NPS55-78-034, An Interactive Software Package For Time Series Analysis, by F.R. Richards and S.R. Woodall, November 1978.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
2.	Department Chairman, Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
3.	Professor P.M. Carrick, Code 54Ca Department of Administrative Science Naval Postgraduate School Monterey, California 93940	2
4.	Professor F. Russell Richards, Code 55Rh Department of Operations Research Naval Postgraduate School Monterey, California 93940	2
5.	Major Paul P. Darling, USMC 8858 Applecross Lane Springfield, Virginia 22153	5
6.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2

Thesis 181684
D1615 Darling
c.1 An analysis and forecast of the supply of first term enlistees to the United States Marine Corps.